Journal of Mathematics and Physics

(Founded by C. L. E. Moore)

Volume XVIII

1939

Editors

FREDERICK G. KEYES
HENRY BAYARD PHILLIPS
DIRK J. STRUIK
MANUEL S. VALLARTA
NORBERT WIENER
FREDERICK S. WOODS

Managing Editor

PHILIP FRANKLIN

OHIO STATE UNIVERSITY

Published by the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

QA1
J87
v.18

OHIO STATE
UNIVERSITY

CONTENTS

Charaktersysteme und Dualitatstheoreme. *By Walther Mayer* 1
A Note on Transcendental Numbers. *By Brockway McMillan* 28
A Complete Set of Asymptotic Formulas for the Whittaker Function and the LaGuerre Polynomials. *By W. C. Taylor* 34
The Light Field. *By A. Gershun* 51
A Function Related to the Series for e^{e^x}. . . . *By L. F. Epstein* 153
The Mutual Interaction of Plasma Electrons. *By W. R. Haseltine* 174
Algebraic Equations with Complex Coefficients. *By F. L. Hitchcock* 202
On Asymptotic Orbits in the Theory of Primary Cosmic Radiation. *By Alfredo Banos, Jr.* 211
Flexure of a Rectangular Covering Slab Elastically Fixed By Its Four Sides. *By S. A. Savin* 239
Synthesis of Reactance 4-Poles Which Produce Prescribed Insertion Loss Characteristics Including Special Applications To Filter Design. *By S. Darlington* 257

550886

CHARAKTERSYSTEME UND DUALITÄTSTHEOREME

VON WALTHER MAYER

Im folgenden wird zu einem gegebenen topologischen Gruppensystem $\sum$ und Abelscher Gruppe K das Σ^K Charakter-System definiert, seine Theorie entwickelt und die Ergebnisse angewandt, hinreichende Bedingungen anzugeben, unter denen die Dualitätstheoreme der kombinatorischen Topologie in der Pontrjaginschen Fassung gelten.

Bezügliche Literatur

L. Pontrjagin: Topol. Dualitatssätze, Mathem. Annalen, Bd. 105, 1931; Theory of Topol. Groups, Annals of Math., Vol. 35, 1934; General Theorem of Duality, Annals of Math., Vol. 35, 1934.

J. W. Alexander: On the Charakters of Discrete Abel. Groups, Annals of Math., Vol. 35, 1934.

J. W. Alexander and L. Zippin: Discrete Groups and their Charakter Groups, Annals of Math., Vol. 36, 1935.

N. E. Steenrod: Universal Homology Groups, Americ. Journal of Math., Vol. 58, 1936.

A. Kolmogoroff: Über die Dualität im Aufbau der komb. Topologie, Recueil Mathem., 1936.

H. Whitney: On Matrices of Integers, Duke Journal, Vol. 3, 1937.

W. Mayer: Topologische Gruppensysteme, erscheint gleichzeitig in den Monatsheften fur Math. und Physik, Wien.

§1. Die K-Charaktergruppe der Gruppe $\mathfrak{A}$

K and $\mathfrak{A}$ seien Abelsche Gruppen. Unter einem K-Charakter von $\mathfrak{A}$ versteht man einen Homomorphismus von $\mathfrak{A}$ in K, d.h. eine in den Elementen a von $\mathfrak{A}$ definierte lineare Funktion $f(\ )$, so dass $f(a) \subset K$.

Sind f und g zwei K-Charaktere von $\mathfrak{A}$, so ist $f + g$ jene in den Elementen von $\mathfrak{A}$ definierte Funktion, für die $(f + g)(a) = f(a) + g(a)$ ist. Offenbar ist $f + g$ selbst ein K-Charakter von $\mathfrak{A}$ und die K-Charaktere von $\mathfrak{A}$ bilden in bezug auf die soeben definierte Summenbildung eine Abelsche Gruppe, die wir die *K-Charaktergruppe* von $\mathfrak{A}$ nennen und mit $\mathfrak{A}^K$ bezeichnen. ($f(a)$ = Guppen-Null von K für jedes $a \subset \mathfrak{A}$ charakterisiert das Nullelement von $\mathfrak{A}^K$).

COLE MEMORIAL LIBRARY

Bemerkung a) Sind $\mathfrak{A}$ und K topologische Gruppen (mit definierten Limesfolgen[1] im Sinne Fréchets oder mit Hausdorf-Umgebungen), so hat es einen Sinn von stetigen K-Charaktern zu reden. Ist K zudem eine kontinuierliche Gruppe ($-f$ ist eine stetige Funktion von f, $f + g$ ist simultan stetig in f und g, $f, g \subset K$), so bilden die *stetigen* K-Charaktere von $\mathfrak{A}$ eine Untergruppe von $\mathfrak{A}^K$.

Ist $\mathfrak{A}_1$ eine Untergruppe von $\mathfrak{A}$, $\mathfrak{A}_1 \subset \mathfrak{A}$, so induziert jeder K-Charakter f von $\mathfrak{A}$ als (lineare) Funktion in den Elementen von $\mathfrak{A}_1$ einen K-Charakter von $\mathfrak{A}_1$.

Die so erklärte Zuordnung $\mathfrak{A}^K \to \mathfrak{A}_1^K$ ist offenbar eine Homomorphie. Ihre *Bildgruppe* enthält alle K-Charaktere von $\mathfrak{A}_1$, die zu K-Charakteren von $\mathfrak{A}$ erweiterbar sind; mit $(\mathfrak{A}_1^K, \mathfrak{A})$ sei diese Gruppe bezeichnet.

Ihr *Kern* besteht aus allen K-Charakteren von $\mathfrak{A}$, durch die Elemente von $\mathfrak{A}_1$ in die Gruppen-Null von K abgebildet werden; mit $(\mathfrak{A}^K, \mathfrak{A}_1)$ bezeichnen wir diese Gruppe. Demnach gilt:

$$(1) \qquad \underline{\mathfrak{A}^K - (\mathfrak{A}^K, \mathfrak{A}_1) \simeq (\mathfrak{A}_1^K, \mathfrak{A}).}$$

Bemerkung b) Sind $\mathfrak{A}$ und K topologische Gruppen und K zudem kontinuierlich, so gilt das zur Isomorphie (1) führende Argument wörtlich für die respektiven Gruppen der stetigen K-Charaktere.

(Es sei dazu bemerkt, dass die in $\mathfrak{A}$ definierte Topologie eine in $\mathfrak{A}_1 (\subset \mathfrak{A})$ induziert: eine Folge $a_1, a_2, \cdots$ ist eine Limesfolge (in $\mathfrak{A}_1$) und a (in $\mathfrak{A}_1$) ihr Limes, wenn diese Aussage für die Folge $a_1, a_2, \cdots$ und das Element a *als Elemente von* $\mathfrak{A}$ gilt)

Wir bezeichnen mit $\{a\}_{\mathfrak{A}_1}$ die das Element a von $\mathfrak{A}$ enthaltende Klasse von $\mathfrak{A} - \mathfrak{A}_1$. Das Element a heisst *ein Repräsentant* von $\{a\}_{\mathfrak{A}_1}$. Ist $f \subset (\mathfrak{A}^K, \mathfrak{A}_1)$, so wird durch

$$(2) \qquad F(\{a\}_{\mathfrak{A}_1}) = f(a)$$

ein K-Charakter F von $\mathfrak{A} - \mathfrak{A}_1$ definiert.

Da der K-Charakter f aus $(\mathfrak{A}^K, \mathfrak{A}_1)$ und der K-Charakter F von $\mathfrak{A} - \mathfrak{A}_1$ sich in (2) ein-eindeutig entsprechen, so definiert (2) die Isomorphie

$$(3) \qquad \underline{(\mathfrak{A} - \mathfrak{A}_1)^K \simeq (\mathfrak{A}^K, \mathfrak{A}_1).}$$

[1] Gewisse unendliche Folgen $a_1, a_2, \cdots$ von Elementen von $\mathfrak{A}$ heissen Limesfolgen und definieren ein in $\mathfrak{A}$ gelegenes Element a, den Limes von $a_1, a_2, \cdots$ ($a = \lim a_n$). Dabei gelte: α) Jede unendliche Teilfolge einer Limesfolge ist selbst eine Limesfolge und definiert dasselbe Limeselement, β) Die Folge $a, a, a, \cdots$ ist eine Limesfolge mit a als Limes.

Bemerkung c) Ist $\mathfrak{A}$ eine kontinuierliche Gruppe und $\mathfrak{A}_1$ eine *abgeschlossene* Untergruppe von $\mathfrak{A}$, so induziert die $\mathfrak{A}$-Topologie eine Topologie in $\mathfrak{A} - \mathfrak{A}_1$:

Es gilt definitionsgemäss

$$(\alpha) \qquad \lim \{a_n\}_{\mathfrak{A}_1} = \{a\}_{\mathfrak{A}_1},$$

wenn für eine geeignet gewählte Repräsentantenfolge a_n der resp. Klassen $\{a_n\}_{\mathfrak{A}_1}$ gilt:

$$(\alpha') \qquad \lim a_n = a + h, \qquad h \subset \mathfrak{A}_1.$$

Damit ist in der Tat die Klasse rechts in (α) bestimmt, d.h. unabhängig von der Auswahl der Repräsentantenfolge. Ist a_n' eine zweite solche Folge, gilt also $a_n - a_n' \subset \mathfrak{A}_1$ und $\lim a_n' = a' + h'$, $h' \subset \mathfrak{A}_1$, so folgt wegen $\mathfrak{A}$ kontinuierlich,

$$(\alpha'') \quad \lim (a_n - a_n') = \lim a_n - \lim a_n' = (a - a') + (h - h').$$

Aber $a_n - a_n'$ ist eine Limes Folge aus $\mathfrak{A}_1$, und da $\mathfrak{A}_1$ abgeschlossen ist, gilt $\lim (a_n - a_n') \subset \mathfrak{A}_1$. Da auch $h - h' \subset \mathfrak{A}_1$, so folgt $a - a' \subset \mathfrak{A}_1$ aus (α''), q.d.e.

Da in bezug auf die in $\mathfrak{A} - \mathfrak{A}_1$ resp. in $\mathfrak{A}$ definierten Topologien die in (2) definierten Funktionen *zugleich* stetig oder unstetig sind, so gilt (3) auch bei Einschränkung auf die resp. Untergruppen der *stetigen* Charaktere. (K ist dabei natürlich eine kontinuierliche Gruppe)

Gilt

$$(4) \qquad \mathfrak{A} \supset \mathfrak{A}_1 \supset \mathfrak{A}_2$$

für drei Gruppen $\mathfrak{A}$, $\mathfrak{A}_1$ und $\mathfrak{A}_2$, so folgt

$$(5) \qquad \mathfrak{A} - \mathfrak{A}_2 \supset \mathfrak{A}_1 - \mathfrak{A}_2,$$

und nach Formel (1) gilt

$$(6) \qquad \underline{|(\mathfrak{A} - \mathfrak{A}_2)^K - ((\mathfrak{A} - \mathfrak{A}_2)^K, \mathfrak{A}_1 - \mathfrak{A}_2) \approx ((\mathfrak{A}_1 - \mathfrak{A}_2)^K, \mathfrak{A} - \mathfrak{A}_2).}$$

In der Isomorphie (3)

$$(7) \qquad (\mathfrak{A} - \mathfrak{A}_2)^K \approx (\mathfrak{A}^K, \mathfrak{A}_2)$$

entspricht der Untergruppe $((\mathfrak{A} - \mathfrak{A}_2)^K, \mathfrak{A}_1 - \mathfrak{A}_2)$ von $(\mathfrak{A} - \mathfrak{A}_2)^K$ die Untergruppe $(\mathfrak{A}^K, \mathfrak{A}_1)$ von $(\mathfrak{A}^K, \mathfrak{A}_2)$.

Beweis: $F(\{a\}_{\mathfrak{A}_2}) \subset ((\mathfrak{A} - \mathfrak{A}_2)^K, \mathfrak{A}_1 - \mathfrak{A}_2)$, wenn $F(\{a\}_{\mathfrak{A}_2}) = 0$, für $a \subset \mathfrak{A}_1$. Damit gilt für die F entsprechende Funktion $f \subset (\mathfrak{A}^K, \mathfrak{A}_2)$ in (2): $f(a) = 0$, für $a \subset \mathfrak{A}_1$, d.h. $f \subset (\mathfrak{A}^K, \mathfrak{A}_1)$.

Umgekehrt $f(a) \subset (\mathfrak{A}^K, \mathfrak{A}_1)$, wenn $f(a) = 0$ für $a \subset \mathfrak{A}_1$. Dann gilt für die entsprechende Funktion $F \subset (\mathfrak{A} - \mathfrak{A}_2)^K$: $F(\{a\}_{\mathfrak{A}_2}) = 0$ für $a \subset \mathfrak{A}_1$, d.h. $F \subset ((\mathfrak{A} - \mathfrak{A}_2)^K, \mathfrak{A}_1 - \mathfrak{A}_2)$.

Somit haben wir

$$(8) \quad (\mathfrak{A} - \mathfrak{A}_2)^K - ((\mathfrak{A} - \mathfrak{A}_2)^K, \mathfrak{A}_1 - \mathfrak{A}_2) = (\mathfrak{A}^K, \mathfrak{A}_2) - (\mathfrak{A}^K, \mathfrak{A}_1)$$

und (6) wird zu

$$(9) \quad \underline{|(\mathfrak{A}^K, \mathfrak{A}_2) - (\mathfrak{A}^K, \mathfrak{A}_1) \approx ((\mathfrak{A}_1 - \mathfrak{A}_2)^K, \mathfrak{A} - \mathfrak{A}_2)}$$

Bemerkung d) *$\mathfrak{A}$ und K seien kontinuierlich und $\mathfrak{A}_2$ abgeschlossen in $\mathfrak{A}$.* Dann gilt (6) für die *stetigen* K-charactere von $\mathfrak{A} - \mathfrak{A}_2$ resp. $\mathfrak{A}_1 - \mathfrak{A}_2$ und in der Isomorphie (7) entsprechen diesen stetigen Charakteren von $\mathfrak{A} - \mathfrak{A}_2$ die stetigen Charaktere aus $(\mathfrak{A}^K, \mathfrak{A}_2)$ (siehe Bemerkung (c)).

Also gilt (9) auch bei Einschränkung auf die respektiven Untergruppen der *stetigen* Charaktere.

§2. Das (n, K)-Charaktersystem Σ^K des topologischen Gruppensystems Σ der Dimension $\leq n$

Wir definieren zuerst die Komplexgruppen $L_i(\Sigma^K)$ und hierauf die Randhomomorphien $L_i(\Sigma^K) \to L_{i-1}(\Sigma^K)$.

a) Wir definieren

$$(1) \quad L_i(\Sigma^K) = (L_{n-i}(\Sigma))^K, \qquad i = 0, 1, \cdots n,$$

die weiteren Komplexgruppen sind alle Nullgruppen.

b) *Die Homomorphie $L_i(\Sigma^K) \to L_{i-1}(\Sigma^K)$.*

Definition: Ist $f_i \subset L_i(\Sigma^K)$, so sei sein Rand

$$(2) \quad R(f_i) = f_i(R(\ )),$$

wo R der Randoperator von Σ ist.

Wir zeigen, dass $f_i(R(\ )) \subset L_{i-1}(\Sigma^K)$, d.h. ein K-Charakter von $L_{n-i+1}(\Sigma)$ ist.

In der Tat, ist $a^{n-i+1} \subset L_{n-i+1}(\Sigma)$, so ist $R(a^{n-i+1}) \subset L_{n-i}(\Sigma)$ und $f_i(R(a^{n-i+1}))$ ist wohldefiniert.

Da weiter $R(\ )$ und $f_i(\ )$ lineare Funktionen sind ihrer respektiven Argumente, so ist $f_i(R(\ ))$ ein K-Charakter von $L_{n-i+1}(\Sigma)$, q.d.e.

c) $R(R(f_i)) = R(f_i(R(\ ))) = f_i(R(R(\ )))$, und da $R(R(a^{n-i+2})) = 0^{n-i}$, so ist $R(R(f_i))$ das Nullelement von $L_{i-2}(\Sigma^K)$. Damit ist Σ^K ein topologisches Gruppensystem: das (n, K)-Charaktersystem von Σ.

Bemerkung a) Wir setzen voraus:

K sei eine kontinuierliche Gruppe, die $L_i(\Sigma)$, $i = 0, 1, \cdots$ topologische Gruppen. Durch $R(H^i)$ sei $L_i(\Sigma)$ stetig in $L_{n-1}(\Sigma)$ abgebildet:

Dann bilden die *stetigen* K-Charaktere ein topologisches Untersystem von Σ^K.

Beweis: Zuerst bilden die stetigen K-Charaktere von $L_i(\Sigma)$ eine Untergruppe aller K-Charaktere von $L_i(\Sigma)$, (K kontinuierliche Gruppe).

Dann ist $R(f_i) \subset L_{i-1}(\Sigma^K)$, wegen $R(\ )$ *stetig*, ($Rf_i = f_i(R(\ ))$), stetig, wenn f_i stetig ist.

Diese beiden Tatsachen aber beweisen die Behauptung.

Bemerkung b) *Sind (wie unter (a)) die $L_i(\Sigma)$ topologische Gruppen und ist R ein stetiger Operator, so ist $Z_i(\Sigma)$ in $L_i(\Sigma)$ abgeschlossen.* ($Z_i(\Sigma)$, Gruppe der i-dimens. Zykeln von Σ.) *Ist zudem noch $L_i(\Sigma)$, $i = 0, 1, \cdots$, kompakt, dann ist auch die i-dimens. Gruppe der berandenden Zykeln, $H_i(\Sigma)$, in $L_i(\Sigma)$ abgeschlossen.*

Beweis: 1) $a^i \subset Z_i(\Sigma)$, wenn $R(a^i) = 0^{i-1}$. Ist $\lim a_n^i = a^i$ eine Limesfolge und $a_n^i \subset Z_i(\Sigma)$, $n = 1, 2, \cdots$, so folgt ($R(a_n^i) = 0^{i-1}$)

$$0^{i-1} = \lim R(a_n^i) = R(\lim a_n^i) = R(a^i), \qquad d.h.\ a^i \subset Z^i(\Sigma).$$

2) $a^i \subset H_i(\Sigma)$, wenn $a^i = R(a^{i+1})$, für ein $a^{i+1} \subset L_{i+1}(\Sigma)$. Ist $\lim a_n^i = a^i$ eine Limesfolge und $a_n^i \subset H_i(\Sigma)$, $n = 1, 2, \cdots$, so folgt $R(a_n^{i+1}) = a_n^i$ für eine geeignete Folge $a_n^{i+1} \subset L_{i+1}(\Sigma)$. Da aber $L_{i+1}(\Sigma)$ kompakt ist, so enthalt die obige Folge eine konvergierende Teilfolge $\lim a_{n'}^{i+1} = a^{i+1}$. Also gilt

$$R(a^{i+1}) = \lim R(a_{n'}^{i+1}) = \lim a_{n'}^i = a^i, \qquad d.h.\ a^i \subset H_i(\Sigma)$$

(In der obigen Überlegung wurde nicht benutzt, dass die topol. Räume $L_i(\Sigma)$, $Z_i(\Sigma)$, $H_i(\Sigma)$ Gruppen sind.)

§3. Relationen zwischen den in Σ und in Σ^K definierten Gruppen

A) *Die Gruppe $Z_i(\Sigma^K)$ der i-dimens. Zykeln von Σ^K.*

$f_i \subset L_i(\Sigma^K) = (L_{n-i}(\Sigma))^K$ ist ein i-dimens. Zyklus, d.h. Element von $Z_i(\Sigma^K)$, wenn $R(f_i) = f_i(R(\ ))$ das Nullelement von $L_{i-1}(\Sigma^K) = (L_{n-i+1}(\Sigma))^K$ ist, das ist wenn

$f_i(R(a^{n-i+1})) = 0$ (Gruppen-Null von K) für *alle* $a^{n-i+1} \subset L_{n-i+1}(\Sigma)$. In anderen Worten: $f_i \subset Z_i(\Sigma^K)$, wenn f_i alle berandenden Zykeln $R(a^{n-i+1})$, d.h. alle Elemente der Untergruppe $H_{n-i}(\Sigma)$ von $L_{n-i}(\Sigma)$ in die Gruppen-Null von K abbildet. In Formel:

$$Z_i(\Sigma) = (L_i(\Sigma^K), H_{n-i}(\Sigma)). \tag{1}$$

Setzt man, §1, (1), $\mathfrak{A} = L_{n-i}(\Sigma)$, $\mathfrak{A}_1 = H_{n-i}(\Sigma)$, so erhält man

$$L_i(\Sigma^K) - Z_i(\Sigma^K) \approx ((H_{n-i}(\Sigma))^K, L_{n-i}(\Sigma)), \tag{2}$$

resp., da $L_i - Z_i \approx H_{i-1}$, für jedes topol. Gruppensystem ist

$$H_{i-1}(\Sigma^K) \approx ((H_{n-i}(\Sigma))^K, L_{n-i}(\Sigma)). \tag{3}$$

Die Formel (3) des §1 gibt

$$Z_i(\Sigma^K) \approx (L_{n-i}(\Sigma) - H_{n-i}(\Sigma))^K. \tag{4}$$

Bemerkung: Sind K und die $L_i(\Sigma)$ kontinuierliche Gruppen und die letzteren auch kompakt, so gelten die obigen Formeln für die resp. Gruppen der kontinuierlichen Charaktere

B) *Die Gruppe $H_i(\Sigma^K)$ der berandenden i-dimens. Zykeln von Σ^K.*

$R(f_{i+1})$ mit $f_{i+1} \subset L_{i+1}(\Sigma^K)$ ist das allgemeine Element von $H_i(\Sigma^K)$. Wegen $R(f_{i+1}) = f_{i+1}(R(\ ))$ folgt, dass $(Rf_{i+1})(a^{n-i}) = f_{i+1}(R(a^{n-i})) = 0$, sobald $a^{n-i} \subset Z_{n-i}(\Sigma)$.

D.h.: Ein Element von $H_i(\Sigma^K)$ bildet alle $n - i$ dimens. Zykeln von Σ, d.h. alle Elemente von $Z_{n-i}(\Sigma)$, in die Gruppen Null von K ab.

Bezeichnet

$$N_i(\Sigma^K) = (L_i(\Sigma^K), Z_{n-i}(\Sigma)), \tag{5}$$

so gilt demnach $N_i(\Sigma^K) \supset H_i(\Sigma^K)$.

Wegen $Z_{n-i}(\Sigma) \supset H_{n-i}(\Sigma)$ folgt aus (1) und (5) $Z_i(\Sigma^K) \supset N_i(\Sigma^K)$. Wir haben demnach:

$$Z_i(\Sigma^K) \supset N_i(\Sigma^K) \supset H_i(\Sigma^K), \tag{6}$$

d.h. $N_i(\Sigma^K)$ ist eine neue Gruppe, die in $Z_i(\Sigma^K)$ liegt und $H_i(\Sigma^K)$ enthält.

Setzen wir in (1) resp. (3) des §1: $\mathfrak{A} = L_{n-i}(\Sigma)$, $\mathfrak{A}_1 = Z_{n-i}(\Sigma)$, so erhalten wir

$$L_i(\Sigma^K) - N_i(\Sigma^K) \approx ((Z_{n-i}(\Sigma))^K, L_{n-i}(\Sigma)), \tag{7}$$

resp.

$$N_i(\Sigma^K) \approx (L_{n-i}(\Sigma) - Z_{n-i}(\Sigma))^K = (H_{n-i-1}(\Sigma))^K. \tag{8}$$

Aus (5) und (8) haben wir

$$(H_{n-i-1}(\Sigma))^K \approx (L_i(\Sigma^K), Z_{n-i}(\Sigma)). \tag{9}$$

C) Weiter gibt §1, (8) für: $\mathfrak{A} = L_{n-i}(\Sigma)$, $\mathfrak{A}_1 = Z_{n-i}(\Sigma)$, $\mathfrak{A}_2 = H_{n-i}(\Sigma)$ die *fundamentale* Relation:

$$\begin{aligned} &(L_i(\Sigma^K), H_{n-i}(\Sigma)) - (L_i(\Sigma^K), Z_{n-i}(\Sigma)) \\ &\qquad \approx ((B_{n-i}(\Sigma))^K, L_{n-i}(\Sigma) - H_{n-i}(\Sigma)), \end{aligned} \tag{10}$$

wo $B_{n-i}(\Sigma) = Z_{n-i}(\Sigma) - H_{n-i}(\Sigma)$, die $n - i^{\text{te}}$ Bettigruppe von Σ ist.

Wegen (1) und (5) konnen wir (10) schreiben:

$$(11) \qquad \underline{Z_i(\Sigma^K) - N_i(\Sigma^K) \approx ((B_{n-i}(\Sigma))^K, L_{n-i}(\Sigma) - H_{n-i}(\Sigma))},$$

aus welcher Formel wir unsere weiteren Schlüsse ziehen werden.

Bemerkung: *K und die $L_i(\Sigma)$ seien kontinuierliche Gruppen, $R(\ )$ sei stetig und die $L_i(\Sigma)$ kompakt.*

Dann ist $H_{n-i}(\Sigma)$ in $L_{n-i}(\Sigma)$ abgeschlossen und aus der Bemerkung (d) des §1 folgt, dass (10) auch für die respektiven Untergruppen der *stetigen Charaktere* gilt.

Damit ist (11) *auch richtig, wenn Σ^K aus dem System der stetigen Charaktere von Σ aufgebaut ist.*

D) Der Vergleich von (3) (für $i + 1$ statt i) und (8) lässt vermuten

$$(12) \quad \underline{N_i(\Sigma^K) - H_i(\Sigma^K) \approx (H_{n-i-1}(\Sigma))^K - ((H_{n-i-1}(\Sigma))^K, L_{n-i-1}(\Sigma))}$$

Diese wichtige Relation gilt tatsächlich, wie im Folgenden gezeigt wird.

Durch die Zuordnung:

$$(13) \qquad \{a^{n-i}\}_{Z_{n-i}(\Sigma)} \leftrightarrow R(a^{n-i}), \qquad a^{n-i} \subset L_{n-i}(\Sigma),$$

wird die Isomorphie hergestellt

$$(14) \qquad L_{n-i}(\Sigma) - Z_{n-i}(\Sigma) \approx H_{n-i-1}(\Sigma).$$

Somit gilt

$$(15) \qquad (L_{n-i}(\Sigma) - Z_{n-i}(\Sigma))^K \approx (H_{n-i-1}(\Sigma))^K$$

infolge der Zuordnung

$$(16) \qquad F(\{a^{n-i}\}_{Z_{n-i}(\Sigma)}) = F(R(a^{n-i}))$$

der respektiven K-Charactere. Da weiter (§1)

$$(17) \quad (L_{n-i}(\Sigma) - Z_{n-i}(\Sigma))^K \approx ((L_{n-i}(\Sigma))^K, Z_{n-i}(\Sigma)) = N_i(\Sigma^K)$$

in bezug auf die Zuordnung (§1, (2),)

$$(18) \quad F(\{a^{n-i}\}_{Z_{n-i}(\Sigma)}) = f(a^{n-i}), \ (f(a^{n-i}) = 0, \text{ wenn } a^{n-i} \subset Z_{n-i}(\Sigma)),$$

ist, so besteht die Isomorphie

$$(19) \qquad (H_{n-i-1}(\Sigma))^K \approx N_i(\Sigma^K)$$

infolge der Zuordnung

$$(20) \qquad F(R(a^{n-i})) = f(a^{n-i}).$$

Wir behaupten nun, dass sich in der Isomorphie (19). (20) die respektiven Untergruppen:

$((H_{n-i-1}(\Sigma))^K, L_{n-i-1}(\Sigma))$ und $H_i(\Sigma^K)$ entsprechen, was dann (12) beweist. Sei

a) $F \subset ((H_{n-i-1}(\Sigma))^K, L_{n-i-1}(\Sigma))$, d.h. es gibt ein $F_{i+1} \subset (L_{n-i-1}(\Sigma))^K$ für das $F_{i+1}(R(a^{n-i})) = F(R(a^{n-i}))$ richtig ist. Für den durch (20) zugeordneten K-Charakter $f \subset N_i(\Sigma^K)$ gilt damit

$$f(a^{n-i}) = F(R(a^{n-i})) = F_{i+1}(R(a^{n-i})) = (R(F_{i+1}))(a^{n-i}) \subset H_i(\Sigma^K).$$

b) $f(a^{n-i}) \subset H_i(\Sigma^K)$ bedeutet $f = F_{i+1}(R(\ ))$ für ein $F_{i+1} \subset (L_{n-i-1}(\Sigma))^K$. Für die durch (20) zugeordnete Funktion $F \subset (H_{n-i-1}(\Sigma))^K$ gilt also

$$F(R(a^{n-i})) = F_{i+1}(R(a^{n-i})),$$

und somit ist $F \subset (H_{n-i-1}(\Sigma))^K$ zu einem K-Charakter $F_{i+1} \subset (L_{n-i-1}(\Sigma))^K$ erweiterbar.

Damit ist (12) *bewiesen.*

Bemerkung: Die Relation (12) ist gleichfalls richtig, wenn nur die respektiven Gruppen der *stetigen* K-Charaktere betrachtet werden.

§4. Hinreichende Bedingungen für $B_i(\Sigma^K) \approx (B_{n-i}(\Sigma))^K$

Vorbemerkung: *Im Falle der stetigen K-Charaktere ist immer vorausgesetzt, dass K und die* $L_i(\Sigma)$ *kontinuierliche Gruppen und die letzteren zudem kompakt seien.* $R(\ )$ *ist ein stetiger Operator.*

Dann folgt aus §3 (11): Für

$$(I) \qquad B_i(\Sigma^K) \approx (B_{n-i}(\Sigma))^K$$

ist das Erfülltsein der beiden Bedingungen

$$(Ia) \qquad (B_{n-i}(\Sigma))^K \approx ((B_{n-i}(\Sigma))^K, L_{n-i}(\Sigma) - H_{n-i}(\Sigma))$$

und

$$(Ib) \qquad N_i(\Sigma^K) = H_i(\Sigma^K)$$

hinreichend. Wir notieren dass (Ia) eine Folge ist von (Ia)'

(Ia)': *Jeder* (event. stetige) *K-Charakter von* $Z_{n-i}(\Sigma)$ *ist zu einem solchen von* $L_{n-i}(\Sigma)$ *erweiterbar.*

In der Tat gilt das dann auch für die speziellen K-Charaktere von $Z_{n-i}(\Sigma)$, die in $H_{n-i}(\Sigma)$ "verschwinden". Diese sind zu Charakteren von $L_{n-i}(\Sigma)$ erweiterbar, die in $H_{n-i}(\Sigma)$ Null sind. Das aber besagt, dass jeder K-Charakter von $B_{n-i}(\Sigma)$ zu einem von $L_{n-i}(\Sigma) - H_{n-i}(\Sigma)$ erweiterbar ist und (Ia) gilt,—sogar mit dem Gleichheitszeichen.

Dagegen ist (Ib) nicht nur Folge sondern äquivalent (§3, (12)):

(Ib'): *Jeder* (event. stetige) *K-Charakter von* $H_{n-i}(\Sigma)$ *ist zu einem solchen von* $L_{n-i}(\Sigma)$ *erweiterbar.*

(Ia)' kann weiter ersetzt werden durch

(Ia)'': $L_{n-i}(\Sigma)$ *ist eine direkte Summe mit* $Z_{n-i}(\Sigma)$ *als Summand:*

$$L_{n-i}(\Sigma) = Z_{n-i}(\Sigma) + A_{n-i}(\Sigma). \tag{1}$$

Dann folgt in der Tat (Ia)'.

Ist nämlich $'f_i \subset (Z_{n-i}(\Sigma))^K$, so definiert

$$f_i(K^{n-i}) = {}'f_i(\underset{-}{K}^{n-i}), \text{ wo} \tag{2}$$

$$K^{n-i} = \underset{-}{K}^{n-i} + \underset{=}{K}^{n-i}, \quad \underset{-}{K}^{n-i} \subset Z_{n-i}(\Sigma), \quad \underset{=}{K}^{n-i} \subset A_{n-i}(\Sigma), \tag{3}$$

die Zerlegung von $K^{n-i} \subset L_{n-i}(\Sigma)$ nach (2) ist, einen K charakter f_i von $L_{n-i}(\Sigma)$ der mit $'f_i$ in $Z_{n-i}(\Sigma)$ übereinstimmt.

(Offenbar kann (Ib)' durch eine (Ia)'' analoge Bedingung (Ib)'' ersetzt werden.)

Endresultat: *Sobald eine der Relationen* (Ia) *mit einer der Relationen* (Ib) *zusammen gilt, besteht* (I).

§5. Σ ist ein freies topol. endliches System der Dimension $\leq n$, K ist die Kreisgruppe (Gruppe der reellen Zahlen modulo 1)

Bekanntlich gilt (Ia)'' des §4. Zum Beweis von

$$B_i(\Sigma^K) \approx (B_{n-i}(\Sigma))^K \tag{1}$$

ist also hinreichend z.B. $N_i(\Sigma^K) = H_i(\Sigma^K)$ zu zeigen.

Ist

$$a^{n-i} = \sum x_\alpha a_\alpha^{n-i}, \qquad \alpha = 1, \cdots, \rho_{n-i}, \tag{2}$$

ein Element von $L_{n-i}(\Sigma)$, dargestellt durch die Basis a_α^{n-i}, $\alpha = 1, \cdots \rho_{n-i}$, so nennen wir die ganzzahligen x_α, $\alpha = 1, \cdots \rho_{n-i}$, die *Darstellungszahlen* von a^{n-i} in bezug auf die Basis a_α^{n-i}. Ist

$$f_i \subset L_i(\Sigma^K) = (L_{n-i}(\Sigma))^K \tag{3}$$

ein K-Charakter von $L_{n-i}(\Sigma)$, so gilt

$$f_i(a^{n-i}) = \sum x_\alpha f_i(a_\alpha^{n-i}) = \sum x_\alpha y_\alpha, \qquad y_\alpha = f_i(a_\alpha^{n-i}) \subset K. \tag{4}$$

(4) bedeutet, dass nach Wahl einer Basis von $L_{n-i}(\Sigma)$ jeder K-Charakter von $L_{n-i}(\Sigma)$ zu einer endlichen Linearform in den Darstellungszahlen x_α der Elemente von $L_{n-i}(\Sigma)$ wird, deren in K liegenden Koeffizienten $y_\alpha = f_i(a_\alpha^{n-i})$ (da $L_{n-i}(\Sigma)$ frei ist) *keiner Einschränkung* unterliegen.

d.h. wenn

$$(6'') \qquad \sum_{\alpha=\sigma_{n-i}+1}^{\rho_{n-i}} y_\alpha x_\alpha = 0 \quad \text{ist, (Gruppen Null von } K)$$

für beliebige x_α, $\alpha = \sigma_{n-i} + 1, \cdots \rho_{n-i}$.

Da K die Kreisgruppe ist, sind (5') und (6'') äquivalente Relationen. In der Tat gibt es zu jeden $y \subset K$ und ganzzahligem E ein $y' \subset K$, so dass $y = Ey'$. Damit sind $y_\alpha = 0$, $\alpha = \sigma_{n-i} + 1, \cdots \rho_{n-i}$ die einzigen Relationen für $\sum y_\alpha x_\alpha \subset H_i(\Sigma^K)$, aber genau dieselben Relationen liefert offenbar (b'').

Damit ist $N_i(\Sigma^K) = H_i(\Sigma^K)$, *und somit* (1) *bewiesen.*

Bemerkung: Von K wurde nur verwendet: Zu jedem $y \subset K$ und ganzzahl. E gibt es ein $y' \subset K$, so dass $y = Ey'$; d.h. $EK = K$. (EK bedeutet die Untergruppe aller Elemente von K der Form Ek, $k \subset K$.)

Somit gilt (1) *auchfür jede Gruppe* K, *für die* $mK = K$ (*für jedes ganzzahl.* m) *richtig ist.*

§6. Das endliche System $\Sigma_{(Y)}$ mit der Koeffizientengruppe Y

Wir gewinnen $\Sigma_{(Y)}$ aus dem *freien* endlichen System Σ durch Einführung der gegebenen Gruppe Y statt der Gruppe G (Gruppe der ganzen Zahlen) *als Koeffizientengruppe* im folgenden Sinn:

Hat $L_i(\Sigma)$ die Basis a_α^i, $\alpha = 1, \cdots \rho_i$, so definieren wir $L_i(\Sigma_{(Y)})$ als Inbegriff der Elemente

$$(1) \qquad K^i = \sum_{\alpha=1}^{\rho_i} y_\alpha a_\alpha^i, \qquad y_\alpha \subset Y,$$

in bezug auf die Addition

$$(1') \qquad \sum y_\alpha a_\alpha^i + \sum y'_\alpha a_\alpha^i = \sum (y_\alpha + y'_\alpha) a_\alpha^i.$$

Weiter bestehe

$$\sum y_\alpha a_\alpha^i = \sum y'_\alpha a_\alpha^i$$

nur als Folge von $y_\alpha = y'_\alpha$, $\alpha = 1, \cdots \rho_i$. Mit $y_\alpha a_\alpha^i$ sei der Komplex K^i (1) bezeichnet, für den nur y_α von Null verschieden ist. (Dann ist (1) eine Summe von ρ_i Komplexen: $y_\alpha a_\alpha^i$, $\alpha = 1, \cdots \rho_i$.)

Die Randhomomorphie $L_i(\Sigma_{(Y)}) \to L_{i-1}(\Sigma_{(Y)})$ wird durch

$$(2) \qquad R(\sum y_\alpha a_\alpha^i) = \sum y_\alpha \epsilon_{\alpha\beta}^i a_\beta^{i-1}, \qquad \alpha = 1, \cdots \rho_i, \beta = 1, \cdots \rho_{i-1},$$

definiert, wo

$$(2') \qquad R(a_\alpha^i) = \sum \epsilon_{\alpha\beta}^i a_\beta^{i-1}$$

(Das bedeutet ferner die Isomorphie der Gruppe $(L_{n-i}(\Sigma))^K$ mit der Gruppe der Linearformen $\Sigma x_\alpha y_\alpha$ infolge der Zuordnung (4), von der wir in diesem § aber keinen Gebrauch machen).

Dies festgestellt suchen wir die Relationen für die Koeffizienten $y_\alpha = f_i(a_\alpha^{n-i})$, damit $f_i \subset H_i(\Sigma^K)$ resp. $f_i \subset N_i(\Sigma^K)$ gilt.

a) $f_i \subset H_i(\Sigma^K)$,

wenn für ein $f_{i+1} \subset L_{i+1}(\Sigma^K)$ gilt: $f_i = R(f_{i+1}) = f_{i+1}(R(\ ))$.

Somit gilt

$$y_\alpha = f_i(a_\alpha^{n-i}) = f_{i+1}(R(a_\alpha^{n-i})) = f_{i+1}(\sum \epsilon_{\alpha\gamma}^{n-i} a_\gamma^{n-i-1}) = \sum_\gamma \epsilon_{\alpha\gamma}^{n-i} f_{i+1}(a_\gamma^{n-i-1}),$$

resp.

$$(5) \qquad y_\alpha = \sum_\gamma \epsilon_{\alpha\gamma}^{n-i} y'_\gamma, \qquad \text{wo} \qquad y'_\gamma = f_{i+1}(a_\gamma^{n-i-1}) \subset K$$

wegen der Willkür des f_{i+1} keiner Einschränkung unterliegt.

D.h. *Damit $f_i \subset H_i(\Sigma^K)$, ist für die Koeffizienten der Linearform $f_i = \sum y_\alpha x_\alpha$ die Relation (5) notwendig und hinreichend.*

b) $f_i \subset N_i(\Sigma^K)$, wenn $f_i(a^{n-i}) = 0$, Gruppen Null von K, als Folge von $R(a^{n-i}) = 0^{n-i-1}$ gilt. Da

$$R(a^{n-i}) = \sum_{\alpha,\gamma} x_\alpha \epsilon_{\alpha\gamma}^{n-i} a_\gamma^{n-i-1},$$

und $L_{n-i-1}(\Sigma)$ eine freie Gruppe ist, so können wir dies auch ausdrucken: $f_i \subset N_i(\Sigma^K)$, wenn $(f_i = \sum y_\alpha x_\alpha)$

$$(6) \qquad \sum y_\alpha x_\alpha = 0, \text{ (Gruppen Null von } K), \textit{ sobald } \sum_\alpha x_\alpha \epsilon_{\alpha\gamma}^{n-i} = 0 \text{ ist.}$$

Für die weitere Diskussion verwenden wir in $L_{n-i}(\Sigma)$, $L_{n-i-1}(\Sigma)$ ein *kanonisches Basispaar*, d.h. eines, in dem die Inzidenz-Matrix, hier $\mathrm{E}_{\alpha\gamma}^{n-i}$ bezeichnet, nur die Diagonalglieder $\mathrm{E}_{\alpha\alpha}^{n-i}$, $\alpha = 1, \cdots \sigma_{n-i}$, von Null verschieden hat. Dabei ist σ_{n-i} der Rang und die $\mathrm{E}_{\alpha\alpha}^{n-i}$ die invarianten Faktoren von $\| \epsilon_{\alpha\gamma}^{n-i} \|$.

Dann gilt $f_i = \sum y_\alpha x_\alpha \subset H_i(\Sigma^K)$, wenn

$$(5') \qquad \begin{cases} y_\alpha = \mathrm{E}_{\alpha\alpha}^{n-i} y'_\alpha, & \alpha = 1, \cdots \sigma_{n-i} \\ y_\alpha = 0, & \alpha = \sigma_{n-i} + 1, \cdots \rho_{n-i}, \text{ und} \end{cases}$$

$$f_i = \sum y_\alpha x_\alpha \subset N_i(\Sigma^K), \text{ wenn}$$

$$(6') \qquad \sum y_\alpha x_\alpha = 0, \quad \text{(Gruppen Null von } K), \text{ sobald}$$

$$x_1 = 0, x_2 = 0, \cdots x_{\sigma_{n-i}} = 0,$$

änderung in Σ, so entsprechen sich die Randelemente (6) und (7) in der Zuordnung (5) für $i-1$ statt i.

Ist Y eine topologische Gruppe (im Fréchet-Sinn) so definieren wir durch:

$$(9) \qquad \lim_{m=\infty} K_m^i = K^i, \quad \text{wo} \quad K_m^i = \sum_\alpha y_{m\alpha} a_\alpha^i, \quad K^i = \sum_\alpha y_\alpha a_\alpha^i$$

Elemente von $L_i(\Sigma_{(Y)})$ sind, *wenn*

$$(9') \qquad \lim_{m=\infty} y_{m\alpha} = y_\alpha, \qquad \alpha = 1, \cdots \rho_i,$$

eine Topologie in $L_i(\Sigma_{(Y)})$, die, *wenn Y noch dazu eine kontinuierliche Gruppe ist, ebenfalls von der Basis Wahl nicht abhängt.*

D.h. dass die $\Sigma_{(Y)} \simeq \Sigma'_{(Y)}$ herstellende Zuordnung (5) zugleich eine topologische (homeomorphe) Abbildung der Räume $L_i(\Sigma_{(Y)})$ und $L_i(\Sigma'_{(Y)})$ ist. Dies folgt aus (4').

Wenn wir Y als *kontinuierliche kompakte* Gruppe voraussetzen, so wird $L_i(\Sigma_{(Y)})$ ebenfalls *kontinuierlich sein und kompakt.*

Der in (2) eingeführte Randoperator bildet dabei $L_i(\Sigma_{(Y)})$ *stetig ab* in $L_{i-1}(\Sigma_{(Y)})$. Ist dann K eine *kontinuierliche Gruppe*, so sind alle Voraussetzungen erfüllt in der (11), §3, folgenden Bemerkung. Wir haben damit das Resultat:

Sind K und Y kontinuierliche Gruppen und Y zudem kompakt, so gilt die Relation (11) *des* §3 *für das aus den stetigen K-Charakteren der $L_i(\Sigma_{(Y)})$ aufgebaute Gruppensystem $(\Sigma_{(Y)})^K$.*

§7. Das K-Charaktersystem $(\Sigma_{(K)})^\kappa$ von $\Sigma_{(K)}$; Σ endlich, K = die Kreisgruppe

Das System $\Sigma_{(K)}$ wurde im §6 definiert:

$$(1) \qquad K^i = \sum y_\alpha a_\alpha^i, \qquad y_\alpha \subset K, \qquad \alpha = 1, \cdots \rho_i,$$

ist das allgemeine Element von $L_i(\Sigma_{(K)})$. $K^i = O^i$ bedeutet $y_\alpha = \theta$, $\alpha = 1, \cdots, \rho_i$.

Wir führen in K, der Gruppe der reellen Zahlen modulo der Gruppe der ganzen Zahlen, die natürliche Topologie ein, die die kontinuierliche Gruppe R aller reellen Zahlen in der Restgruppe $R - G$, G die (in R) abgeschlossene Gruppe der ganzen Zahlen, induziert (Siehe §1 Bemerkung c);

Die Folge $\{y_n\}$, $n = 1, 2, \cdots$, $\{y_n\} \subset K$ konvergiert gegen $\{y\} \subset K$, sobald für geeignete Klassenrepräsentanten der $\{y_n\}$ resp des $\{y\}$

$$(2) \qquad \lim y_n = y$$

(als Limesfolge von R) besteht.

die Randhomomorphie $L_i(\Sigma) \to L_{i-1}(\Sigma)$ bestimmt. Da offenbar $R(R(\sum y_\alpha a_\alpha^i)) = 0^{i-2}$ ist, *so ist* $\Sigma_{(Y)}$, *gegeben durch die Gruppen* $L_i(\Sigma_{(Y)})$ *mit den Randhomomorphien* (2), *ein topologisches Gruppensystem.*

$\Sigma_{(Y)}$ wurde in Bezug auf eine gegebene Basis a_α^i, $i = 0, 1, \cdots$ von Σ definiert; wir zeigen nun, dass $\Sigma_{(Y)}$ aber in Wirklichkeit von der Basiswahl in Σ unabhängig ist.

Sei $a_\alpha'^i$, $\alpha = 1, \cdots \rho_i$, $i = 0, 1, \cdots$ ein zweites Basissystem in Σ, dann gelten bekanntlich die Transformations Relationen

$$a_\alpha^i = \sum p_{\alpha\beta}^i a_\beta'^i, \qquad a_\gamma'^i = \sum \bar{p}_{\gamma\alpha}^i a_\alpha^i, \qquad i = 0, 1, \cdots, \tag{3}$$

mit ganzzahligen Matrizen p, $\bar{p}$, die zueinander invers und also unimodular sind. Der Transformation (3) in Σ lassen wir in $\Sigma_{(Y)}$ eine Transformation entsprechen

$$(ya_\alpha^i) = \sum p_{\alpha\beta}^i (ya_\beta'^i), \qquad (ya_\gamma'^i) = \sum \bar{p}_{\gamma\alpha}^i (ya_\alpha^i), \tag{3'}$$

durch die in $L_i(\Sigma_{(Y)})$ die neue "Basis" $ya_\gamma'^i$, $\gamma = 1, \cdots \rho_i$ eingeführt wird: Jedes Element $\sum y_\alpha a_\alpha^i \subset L_i(\Sigma_{(Y)})$ ist durch die neue "Basis" darstellbar

$$\sum y_\alpha a_\alpha^i = \sum p_{\alpha\beta}^i y_\alpha a_\beta'^i = \sum y_\beta' a_\beta'^i, \qquad \text{wo} \tag{4}$$

$$y_\beta' = \sum p_{\alpha\beta}^i y_\alpha. \tag{4'}$$

Sei nun $\Sigma'_{(Y)}$ das System, das analog $\Sigma_{(Y)}$ aber bei Verwendung der Basis $a_\alpha'^i$ von Σ konstruiert ist. *Wir beweisen*: $\Sigma_{(Y)} = \Sigma'_{(Y)}$. Sei $\sum y_\alpha a_\alpha^i \subset L_i(\Sigma_{(Y)})$, dann können wir nach Einführung der Basis $ya_\alpha'^i$ in $\Sigma_{(Y)}$ nach (4): $\sum y_\alpha a_\alpha^i = \sum y_\beta' a_\beta'^i$ schreiben.

Wir ordnen dann $\sum y_\alpha a_\alpha^i = \sum y_\beta' a_\beta'^i \subset L_i(\Sigma_{(Y)})$ zu das Element $\sum y_\beta' a_\beta' \subset L_i(\Sigma'_{(Y)})$. Diese Zuordnung:

$$\sum y_\alpha a_\alpha^i \leftrightarrow \sum y_\beta' a_\beta'^i \tag{5}$$

ist (wie in dem Doppelpfeil angedeutet) ein—eindeutig und stellt eine Isomorphie $L_i(\Sigma_{(Y)}) \simeq L_i(\Sigma'_{(Y)})$ her, deren Randtreue wir zeigen.

Der Rand des Elementes von $L_i(\Sigma_{(Y)})$ links in (5) ist:

$$R(\sum y_\alpha a_\alpha^i) = \sum y_\alpha \epsilon_{\alpha\sigma}^i a_\sigma^{i-1} = \sum y_\alpha \epsilon_{\alpha\sigma}^i p_{\sigma\tau}^{i-1} a_\tau'^{i-1} \subset L_{i-1}(\Sigma_{(Y)}), \tag{6}$$

der Rand des Elementes von $L_i(\Sigma'_{(Y)})$ rechts in (5) ist:

$$R(\sum y_\beta' a_\beta'^i) = \sum y_\beta' \epsilon_{\beta\tau}'^i a_\tau'^{i-1} = \sum y_\alpha p_{\alpha\beta}^i \epsilon_{\beta\tau}'^i a_\tau'^{i-1} \subset L_{i-1}(\Sigma'_{(Y)}); \tag{7}$$

da aber

$$\sum_\sigma \epsilon_{\alpha\sigma}^i p_{\sigma\tau}^{i-1} = \sum_\beta p_{\alpha\beta}^i \epsilon_{\beta\tau}'^i \tag{8}$$

die bekannte Transformation der Inzidenz Matrizen ist bei Basis-

Bezeichnung: Wir wollen in der Folge die die *Klasse* $\{y\}$ bezeichnende Klammer weglassen $y \subset K$ bedeutet also die Klasse $y + g$, g ganze Zahl.

Damit ist auch (§6) *in* $L_i(\Sigma_{(K)})$ *eine Topologie definiert, in bezug auf die diese Gruppen kontinuierlich und kompakt sind.*

(K ist offenbar kontinuierlich und wie leicht einzusehen auch kompakt.) Wir können daher das Schlussresultst des §6 anwenden und

$$B_i(\Sigma^K_{(K)}) \approx (B_{n-i}(\Sigma_{(K)}))^K \tag{3}$$

beweisen, indem wir z.B. zeigen, dass a) jeder in $Z_i(\Sigma_{(K)})$ definierte *stetige* Charakter zu einem solchen in $L_i(\Sigma_{(K)})$ erweiterbar ist, und dass b) $N_i(\Sigma^K_{(K)}) = H_i(\Sigma^K_{(K)})$ ist.

Bemerkung: (3) gilt sowohl für das allgemeine $\Sigma^K_{(K)}$ als auch für dessen Untersystem, das aus den *stetigen* K-Charakteren aufgebaut ist.

(*Denn jeder* K*-Charakter von* $Z_i(\Sigma_{(K)})$ *ist zu einem von* $L_i(\Sigma_{(K)})$ *erweiterbar, und auch* $N_i(\Sigma^K_{(K)}) = H_i(\Sigma^K_{(K)})$ *gilt ganz allgemein.*)

Wir beschränken uns aber im folgenden auf das System, das aus den *stetigen* K-Charakteren gebildet ist und welches gleichfalls mit $\Sigma^K_{(K)}$ bezeichnet sei.

(Für den allgemeinen Fall verweisen wir auf das Resultat des §7a.)

Wir werden zuerst untersuchen, welche Form ein *stetiger* K-Charakter von $L_i(\Sigma_{(K)})$, $f_{n-i} \subset L_{n-i}(\Sigma^K_{(K)})$, hat.

Neben der Linearitat der Funktion $f_{n-i}(K^i)$ gilt

$$\lim f_{n-i}(K^i_m) = f_{n-i}(\lim K^i_m) \tag{4}$$

für jede konvergente Folge K^i_m aus $L_i(\Sigma_{(K)})$. Ist $K^i \subset L_i(\Sigma_{(K)})$ und (1) seine Darstellung durch die Basis a^i_α, $\alpha = 1, \cdots \rho_i$, so haben wir

$$f_{n-i}(K^i) = \sum_{\alpha} f_{n-i}(y_\alpha a^i_\alpha), \qquad \alpha = 1, \cdots \rho_i. \tag{5}$$

Sei α zwischen 1 und ρ_i gewählt und y_α eine feste *irrationale* Klasse in K, so bestimmt

$$f_{n-i}(y_\alpha a^i_\alpha) = z_\alpha \tag{5'}$$

die Klasse z_α aus K. Zwischen y_α und z_α besteht eine Relation

$$ay_\alpha + bz_\alpha + c = 0 \tag{6}$$

mit ganzzahligen a, b und c. (D.h. $ay_\alpha + bz_\alpha = 0$ als Relation in K.) Sonst gäbe es nach einem Theorem Kroneckers für beliebige Elemente

(Klassen) $\tilde{y}_\alpha$, $\tilde{z}_\alpha$ von K eine ganzzahlige Folge m_ν, $\nu = 1, 2, \cdots \infty$, so dass

$$\lim_{\nu=\infty} m_\nu y_\alpha = \tilde{y}_\alpha, \qquad \lim_{\nu=\infty} m_\nu z_\alpha = \tilde{z}_\alpha \tag{7}$$

zugleich bestehen.

Dann aber führt Multiplikation von (5′) mit m_ν und darauf folgender Grenzübergang $\nu \to \infty$ wegen (4) und (7) auf

$$f_{n-i}(\tilde{y}_\alpha a_\alpha^i) = \tilde{z}_\alpha,$$

einer wegen der Willkur von $\tilde{y}_\alpha$, $\tilde{z}_\alpha$ unmöglichen Relation. *Damit gilt* (6). Multiplikation von (5′) mit b ergibt somit

$$f_{n-i}(by_\alpha a_\alpha^i) = -ay_\alpha \tag{8}$$

als Relation in K.

Da y_α irrational gewählt wurde, und es daher keine Relation $\tilde{a}y + \tilde{b} = 0$ mit ganzzahligen $\tilde{a}$, $\tilde{b}$ gibt, so existiert nach dem eben verwendeten Kroneckerschen Theorem zu dem beliebigen $\tilde{y}_\alpha \subset K$ eine ganzzahlige Folge m_ν, $\nu = 1, 2, \cdots \infty$, so dass gilt

$$\lim_{\nu=\infty} m_\nu y_\alpha = \tilde{y}_\alpha. \tag{9}$$

Offenbar (a und b sind ganze Zahlen) hat (9)

$$\lim m_\nu b y_\alpha = b\tilde{y}_\alpha \quad \text{und} \quad \lim m_\nu a y_\alpha = a\tilde{y}_\alpha \text{ zur Folge.}$$

Aus

$$f_{n-i}(m_\nu b y_\alpha a_\alpha^i) = -m_\nu a y_\alpha \tag{10}$$

ergibt der Grenzübergang für $\nu = \infty$

$$f_{n-i}(b\tilde{y}_\alpha a_\alpha^i) = -a\tilde{y}_\alpha. \tag{10′}$$

Ersetzt man in (10′) $\tilde{y}_\alpha$ durch $\tilde{y}_\alpha + 1/b$, so bleibt die linke und damit auch die rechte Seite dieser Relation in K ungeändert.

Damit ist $-a(\tilde{y}_\alpha + 1/b) + a\tilde{y}_\alpha = -a/b$ ganzzahlig. Bezeichnet x_α diese ganze Zahl, $x_\alpha = -a/b$, und setzt man $y'_\alpha = \tilde{y}_\alpha b$, so gibt (10′)

$$f_{n-i}(y'_\alpha a_\alpha^i) = y'_\alpha x_\alpha, \tag{10″}$$

x_α *ganze Zahl*, für jedes y'_α in K.

(Denn jedes y'_α kann in die Form $\tilde{y}_\alpha b$ gebracht werden.)

Wir haben demnach das Resultat:

Nach Wahl der "Basis a'_α in $L_i(\Sigma_{(K)})$ wird jeder stetige K-Charakter

f_{n-i} zu einer endlichen Linearform in den Darstellungszahlen $y_\alpha \subset K$ der Elemente von $L_i(\Sigma_{(K)})$ in bezug auf diese Basis

$$f_{n-i}(K^i) = \sum x_\alpha y_\alpha, \qquad \alpha = 1, \cdots \rho_i, \tag{11}$$

deren Koeffizienten ganzzahlig sind und offenbar keiner Einschränkung unterliegen. (Eine Form (11) mit belieb. ganzzahl. x_α ist ein *stetiger* K-Charakter von $L_i(\Sigma_{(K)})$).

Dies festgestellt, wollen wir $Z_i(\Sigma_{(K)})$ die Gruppe der i-dimens. Zykeln von $\Sigma_{(K)}$ unter Verwendung kanonischer Basen in $L_i(\Sigma_{(K)})$ resp. $L_{i-1}(\Sigma_{(K)})$ untersuchen. Bezeichnen wir diese Basen mit $a_\alpha'^i$ resp. $a_\beta'^{i-1}$, so folgt

$$\sum y_\alpha a_\alpha'^i \subset Z_i(\Sigma_{(K)}), \quad \text{wenn} \tag{12}$$
$$R(\sum y_\alpha a_\alpha'^i) = \sum y_\alpha \mathrm{E}^i_{\alpha\alpha} a_\alpha'^{i-1} = 0^{i-1},$$

d.h. wenn $y_\alpha \mathrm{E}^i_{\alpha\alpha} = 0$ (Gruppen-Null von K), $\alpha = 1, \cdots \rho_i$.

Dabei sind die ganzzahligen $\mathrm{E}^i_{\alpha\alpha}$ die invar. Faktoren der Inzidenz-Matrix $|| \epsilon^i_{\alpha\beta} ||$ irgend eines Basispaars a^i_α, a^{i-1}_β von $L_i(\Sigma_{(K)})$ resp. $L_{i-1}(\Sigma_{(K)})$. Ist σ_i der Rang dieser Matrix, so sind nur die $\mathrm{E}^i_{\alpha\alpha}$, $\alpha = 1, \cdots \sigma_i$ von Null verschieden

Wir haben also:

$$\sum y_\alpha a_\alpha'^i \subset Z_i(\Sigma_{(K)}), \quad \text{wenn} \quad y_\alpha = g_\alpha/\mathrm{E}^i_{\alpha\alpha}, \tag{13}$$
$$\alpha = 1, \cdots \sigma_i, \text{ mit ganzzahl.}$$

g_α und die restlichen $y_{\sigma_i+1}, \cdots y_{\rho_i}$ beliebig sind.

Damit ist $Z_i(\Sigma_{(K)})$ die direkte Summe von $\rho_i - \sigma_i$ Kreisgruppen: $y_\alpha a_\alpha'^i$, $\alpha = \sigma_i + 1, \cdots \rho_i$ und von σ_i endlichen Zykelngruppen: $y_\alpha a_\alpha'^i$, $\alpha = 1, \cdots \sigma_i$, $y_\alpha = g_\alpha/\mathrm{E}^i_{\alpha\alpha}$, der resp. Ordnungen $\mathrm{E}^i_{\alpha\alpha}$.

Ist nun f_{n-i} ein stetiger K-Charakter von $Z_i(\Sigma_{(K)})$, so ist $f_{n-i}(K^i)$, $K^i \subset Z_i(\Sigma_{(K)})$, die Summe der Anteile, die durch die obigen Summanden geliefert werden.

Wie zuvor gilt $f_{n-i}(y_\alpha a_\alpha'^i) = x_\alpha y_\alpha$, mit ganzzahligem x_α, für $\alpha = \sigma_i + 1, \cdots \rho_i$, als Folge der Stetigkeit von f_{n-i}.

Sei jetzt $\alpha = 1, \cdots \sigma_i$. Wir setzen

$$f_{n-i}(1/\mathrm{E}^i_{\alpha\alpha} a_\alpha'^i) = z_\alpha \subset K. \tag{14}$$

Dann ist $\mathrm{E}^i_{\alpha\alpha} z_\alpha = 0$, die Gruppen Null von K, d.h. $\mathrm{E}^i_{\alpha\alpha} z_\alpha$ ist gleich der ganzen Zahl x_α; $z_\alpha = x_\alpha/\mathrm{E}^i_{\alpha\alpha}$.

Somit gilt für jedes ganzzahl. g_α

$$f_{n-i}(g_\alpha/\mathrm{E}^i_{\alpha\alpha} a_\alpha'^i) = x_\alpha g_\alpha/\mathrm{E}^i_{\alpha\alpha}. \tag{15}$$

Demnach ist für $\alpha = 1, 2, \cdots \rho_i$

$$f_{n-i}(y_\alpha a_\alpha'^i) = x_\alpha y_\alpha \tag{16}$$

mit ganzzahligem x_α richtig, *und* $\sum x_\alpha y_\alpha$, $\alpha = 1, \cdots \rho_i$, *mit diesen* x_α *ist offenbar* ein stetiger *K-Charakter von* $L_i(\Sigma_{(K)})$, *der in* $Z_i(\Sigma_{(K)})$ *mit* f_{n-i} übereinstimmt.

(Wir bemerken, dass (15) das ganzzahlige x_α nur modulo $\mathrm{E}^i_{\alpha\alpha}$ bestimmt, d.h. dass die Erweiterung eines K-Charakters von $Z_i(\Sigma_{(K)})$ zu einem von $L_i(\Sigma_{(K)})$ keineswegs eindeutig ist.)

Sehr einfach ist der Beweis von $H_i(\Sigma^K_{(K)}) = N_i(\Sigma^K_{(K)})$, wenn wir das eben eingeführte kanonische Basispaar verwenden:

a) $\sum x_\alpha y_\alpha \subset N_i(\Sigma^K_{(K)})$, wenn

(17) $\sum x_\alpha y_\alpha = 0$, (Gruppen Null von K), sobald $\sum y_\alpha a_\alpha'^i \subset Z_i(\Sigma_{(K)})$, d.h. sobald (13): $y_\alpha = g_\alpha/\mathrm{E}^i_{\alpha\alpha}$, $\alpha = 1, \cdots \sigma_i$, ($g_\alpha$, ganzzahlig), und die übrigen y_α beliebige Elemente von K sind.

Das ist offenbar dann und nur dann der Fall, wenn

$$\begin{cases} x_\alpha = \mathrm{E}^i_{\alpha\alpha} x'_\alpha, & (x'_\alpha, \text{ ganze Zahl}). \qquad \alpha = 1, \cdots \sigma_i \\ x_\alpha = 0, & \alpha = \sigma_{i+1}, \cdots \rho_i. \end{cases} \tag{18}$$

b) $\sum x_\alpha y_\alpha \subset H_i(\Sigma^K_{(K)})$, wenn $f_{n-i}(K^i) = \sum x_\alpha y_\alpha$ und $f_{n-i} = R(f_{n-i+1}) = f_{n-i+1}(R(\ \ ))$.

Damit gilt ($K^i = \sum y_\alpha a_\alpha'^i$)

$$\begin{cases} f_{n-i}(K^i) = \sum x_\alpha y_\alpha = f_{n-i+1}(R(\sum y_\alpha a_\alpha'^i)) = f_{n-i+1}(\sum y_\alpha \mathrm{E}^i_{\alpha\alpha} a_\alpha'^{i-1}) \\ \qquad = \sum f_{n-i+1}(y_\alpha \mathrm{E}^i_{\alpha\alpha} a_\alpha'^{i-1}) = \sum y_\alpha \mathrm{E}^i_{\alpha\alpha} x'_\alpha. \end{cases} \tag{19}$$

Vergleich von (18) *und* (19) *zeigt* $H_i(\sum {}^K_{(K)}) = N_i(\Sigma^K_{(K)})$.

(In der Tat ist x'_α in (19) ganzzahlig und durch $f_{n-i+1}(y'_\alpha a_\alpha'^{i-1}) = y'_\alpha x'_\alpha$ definiert und ist somit wegen der Willkür des f_{n-i+1} keiner Einschränkung unterworfen.)

§8. Eine Verallgemeinerung der Resultate der Paragraphen 5–7

Wir bezeichnen (Y eine gegebene Gruppe): mit ${}_{\mathrm{E}}Y$ die Untergruppe der Elemente $y \subset Y$, für die $\mathrm{E}y = 0$ ist und mit $\mathrm{E}Y$ die Untergruppe der Elemente der Form $\mathrm{E}y$ von Y, (E ganzeZahl).

Theorem: *Ist* Σ *endlich, so gilt*

$$B_i(\Sigma^K_{(Y)}) = (B_{n-i}(\Sigma_{(Y)}))^K, \tag{I}$$

wenn jeder K-Charakter von ${}_{\mathrm{E}}Y$ *und von* $\mathrm{E}Y$ *zu einem von* Y *erweiterbar ist.*

Wir beweisen (Ia)′ und (Ib)′ des §4 and damit (I)

Beweis von (Ia)′: Wir führen in $L_i(\Sigma)$ und $L_{i-1}(\Sigma)$ ein kanonisches Basispaar ein, in bezug auf das die Randrelationen die spezielle Form haben:

$$R(a_\alpha^i) = \mathrm{E}_{\alpha\alpha}^i a_\alpha^{i-1}, \qquad \alpha = 1, \cdots \rho_i. \tag{1}$$

Damit gilt $R(\sum y_\alpha a_\alpha^i) = \sum y_\alpha \mathrm{E}_{\alpha\alpha}^i a_\alpha^{i-1}$, und $\sum y_\alpha a_\alpha^i \subset Z_i(\Sigma_{(Y)})$, wenn $y_\alpha \mathrm{E}_{\alpha\alpha}^i = 0$ (Gruppen Null von Y). D.h. $\sum y_\alpha a_\alpha^i \subset Z_i(\Sigma_{(Y)})$, wenn $y_\alpha = {}_{\mathrm{E}}\mathrm{E}_{\alpha\alpha}^i Y$.

Damit ist $Z_i(\Sigma_{(Y)})$ isomorph der direkten Summe der ρ_i Gruppen ${}_{\mathrm{E}}\mathrm{E}_{\alpha\alpha}^i Y$, und da jeder K-Charakter eines Summanden dieser Summe zu einem von Y erweiterbar ist, so ist jeder K-Charakter von $Z_i(\Sigma_{(Y)})$ zu einem der direkten Summe der ρ_i Gruppen $y_\alpha a_\alpha^i$ $(\approx Y)$ erweiterbar, d.h. zu einem K-Charakter von $L_i(\Sigma_{(Y)})$.

Beweis von (Ib)′: Wir führen in $L_{i+1}(\Sigma)$ und $L_i(\Sigma)$ ein kanonisches Basispaar ein.

Dann ist $\sum y_\alpha \mathrm{E}_{\alpha\alpha}^{i+1} a_\alpha^i$ das allgemeine Element von $H_i(\Sigma_{(Y)})$ und diese Gruppe ist somit isomorph der direkten Summe der ρ_i Gruppen $\mathrm{E}_{\alpha\alpha}^{i+1} Y$, $\alpha = 1, \cdots \rho_i$. Da aber jeder K-Charakter eines solchen Summanden zu einem von Y erweiterbar ist, so folgt, dass jeder K-Charakter von $H_i(\Sigma_{(Y)})$ zu einem von $L_i(\Sigma_{(Y)})$ erweiterbar ist.

Damit ist (I) bewiesen.

Sind K und Y kontinuierliche Gruppen und Y zudem kompakt, so kann man, (§6), durch

$$\sum y_{m\alpha} a_\alpha^i \to \sum y_\alpha a_\alpha^i, \qquad \text{wenn} \qquad y_{m\alpha} \to y_\alpha, \quad (m \to \infty),$$

eine Topologie in $L_i(\Sigma_{(Y)})$ erklären, in bezug auf die dann diese Gruppen kontinuierlich und kompakt sind. §3 (11) gilt dann auch für das Gruppensystem der *stetigen* Charaktere, (§6).

Somit gilt (I) *auch für das System $\Sigma_{(Y)}^K$ der stetigen K-Charaktere, wenn jeder stetige K-Charakter von ${}_{\mathrm{E}}Y$ und jeder stetige K-Charakter von $\mathrm{E}Y$ zu einem stetigen K-Charakter von Y erweiterbar ist.*

Wir bemerken dazu, dass jeder K-Charakter von $\mathrm{E}Y$ zu einem von Y erweiterbar ist, wenn:

$$(Y^K, {}_{\mathrm{E}}Y) = \mathrm{E}(Y^K). \tag{α}$$

Beweis:

$$Y - {}_{\mathrm{E}}Y \approx \mathrm{E}Y \quad \text{infolge} \quad \{y\}\,{}_{\mathrm{E}}Y \leftrightarrow \mathrm{E}y, \qquad y \subset Y.$$

Folglich ist

$$(Y - {}_{\mathrm{E}}Y)^K \approx (\mathrm{E}Y)^K, \quad \text{infolge} \quad F(\{y\}_{\mathrm{E}}Y) = F(\mathrm{E}y).$$

Im §1 zeigten wir

$$(Y - {}_{\mathrm{E}}Y)^K \approx (Y^K, {}_{\mathrm{E}}Y), \quad \text{infolge} \quad F(\{y\}_{\mathrm{E}}Y) = f(y),$$

also ist

$$(\beta) \qquad (Y^K, {}_{\mathrm{E}}Y) \approx (\mathrm{E}Y)^K, \quad \text{infolge} \quad f(y) = F(\mathrm{E}y),$$

was offenbar auch direkt einleuchtet.

Wegen (α) folgt aus $f \subset (Y^K, {}_{\mathrm{E}}Y)$, dass $f = \mathrm{E}f'$, $f' \subset Y^K$ ist Damit folgt (β) aus $F \subset (\mathrm{E}Y)^K$ dass $F(\mathrm{E}y) = \mathrm{E}f'(y)$.

In $f'(y) \subset Y^K$ haben wir somit einen K-Charakter von Y, der in $\mathrm{E}Y$ mit F übereinstimmt, $F \subset (\mathrm{E}Y)^K$.

§9. Das erste (Poincarésche) Dualitätstheorem in der Pontrjaginschen Fassung

Σ und Σ^*, beide von der Dimension $\leq n$ seien freie, endliche n-duale Systeme im (bereits) klassischem Sinne (in bezug auf ein Basispaar von Σ resp. Σ^*):

a) Zwischen $L_i(\Sigma)$ und $L_{n-i}(\Sigma^*)$ besteht eine Isomorphie, in der die resp. Basiselemente $a_\alpha^i \subset L_i(\Sigma)$ und $a_\alpha^{*n-i} \subset L_{n-i}(\Sigma^*)$, $\alpha = 1, \cdots \rho_i$, einander entsprechen.

Weiter gilt

b) $$\epsilon_{\beta\alpha}^{*n-i} = \epsilon_{\alpha\beta}^{i+1}$$

für die Inzidenzmatrizen in bezug auf die obigen Basen, d.h.

$$R(a_\alpha^{i+1}) = \sum \epsilon_{\alpha\beta}^{i+1} a_\beta^i, \qquad R(a_\beta^{*n-i}) = \sum \epsilon_{\beta\alpha}^{*n-i} a_\alpha^{*n-i-1}.$$

Wir betrachten Σ und $\Sigma_{(K)}^*$, K = Kreisgruppe, und zeigen

(I) $\Sigma^K \approx \Sigma_{(K)}^*$,

und

(II) $(\Sigma_{(K)}^*)^K \approx \Sigma$ (bei Einschränkung auf *stetige* K-Charaktere).

Beweis von (I). Wir zeigen zuerst $L_{n-i}(\Sigma_{(K)}^*) \approx L_{n-i}(\Sigma^K) = (L_i(\Sigma))^K$, und weiter, dass die diese Isomorphie herstellende Zuordnung randtreu ist.

Ist

$$f_{n-i} \subset L_{n-i}(\Sigma^K) \quad \text{und} \quad K^i = \sum x_\alpha a_\alpha^i, \qquad (x_\alpha \text{ ganze Zahlen}),$$

so ist

$$f_{n-i}(K^i) = \sum f_{n-i}(a_\alpha^i)x_\alpha = \sum y_\alpha x_\alpha\,, \qquad y_\alpha = f_{n-i}(a_\alpha^i) \subset K,$$

eine Linearform, der wir in $L_{n-i}(\Sigma^*_{(K)})$ das Element $\sum y_\alpha a_\alpha^{*n-i}$ zuordnen. Diese Zuordnung $L_{n-i}(\Sigma^K) \to L_{n-i}(\Sigma^*_{(K)})$:

$$(\alpha) \qquad f_{n-i} \leftrightarrow \sum f_{n-i}(a_\alpha^i)a_\alpha^{*n-i}$$

ist offenbar eine Isomorphie.

Nun folgt aus (α)

$$(\alpha') \qquad R(f_{n-i}) \leftrightarrow R(\sum f_{n-i}(a_\alpha^i)a_\alpha^{*n-i}) = \sum f_{n-i}(a_\alpha^i)\epsilon_{\alpha\beta}^{*n-i}a_\beta^{*n-i-1},$$

was im folgenden bewiesen wird.

$$Rf_{n-i} = f_{n-i}(R(\ )) \subset L_{n-i-1}(\Sigma^K) = (L_{i+1}(\Sigma))^K.$$

Diesem Element entspricht in der Zuordnung (α) (für $n-i-1$ statt $n-i$) in $L_{n-i-1}(\Sigma^*_{(K)})$:

$$\sum f_{n-i}(R(a_\beta^{i+1}))a_\beta^{*n-i-1} = \sum f_{n-i}(\sum \epsilon_{\beta\alpha}^{i+1}a_\alpha^i)a_\beta^{*n-i-1}$$
$$= \sum f_{n-i}(a_\alpha^i)\epsilon_{\beta\alpha}^{i+1}a_\beta^{*n-i-1} = \sum f_{n-i}(a_\alpha^i)\epsilon_{\alpha\beta}^{*n-i}a_\beta^{*n-i-1}$$

d.h. das Element rechts in (α'), q.d.e.

Beweis von (II).

Wir zeigen zuerst $L_{n-i}(\Sigma) \approx L_{n-i}(\Sigma^{*K}_{(K)}) = (L_i(\Sigma^*_{(K)}))^K$, und weiter die Randtreue der diese Isomorphie herstellenden Zuordnung.

Ist

$$f_{n-i} \subset L_{n-i}(\Sigma^K_{(K)}) \quad \text{und} \quad K^{*i} = \sum y_\alpha a_\alpha^{*i} \subset Li(\Sigma^*_{(K)}),$$

so ordnen wir f_{n-i} vermittels der es reprasentierenden Linearform

$$f_{n-i}(K^{*i}) = \sum f_{n-i}(y_\alpha a_\alpha^{*i}) = \sum x_\alpha y_\alpha\,, \qquad (x_\alpha \text{ ganze Zahlen}),$$

das Element $\sum x_\alpha a_\alpha^{n-i}$ von $L_{n-i}(\Sigma)$ zu. Diese Zuordnung

$$(\beta) \qquad f_{n-i} \leftrightarrow f_{n-i}(K^{*i}) \leftrightarrow \sum x_\alpha a_\alpha^{n-i},$$

die offenbar eine Isomorphie ist, ist randtreu.

$$\text{Sei} \quad K^{*i+1} = \sum a_\beta^{*i+1}y_\beta \subset L_{i+1}(\Sigma^*_{(K)}), \quad \text{dann gilt}$$

$$(Rf_{n-i})(K^{*i+1}) = f_{n-i}(R(K^{*i+1})) = f_{n-i}(\sum_{\alpha,\beta} \epsilon_{\beta\alpha}^{*i+1}a_\alpha^{*i}y_\beta)$$
$$= \sum_\alpha f_{n-i}\Big(\Big(\sum_\beta \epsilon_{\beta\alpha}^{*i+1}y_\beta\Big)a_\alpha^{*i}\Big) = \sum_{\alpha,\beta} y_\beta\epsilon_{\beta\alpha}^{*i+1}x_\alpha.$$

Daher ist Rf_{n-i} in (β) zugeordnet das Element: $\sum_{\alpha,\beta} \epsilon_{\beta\alpha}^{*i+1} x_\alpha a_\beta^{n-i-1} \subset$ $L_{n-i-1}(\Sigma)$, das ist $R(\sum x_\alpha a_\alpha^{n-i})$ wegen (b).

Damit ist die Zuordnung (β) als randtreu erkannt und (II) bewiesen.

Da wir früher $(B_{n-i}(\Sigma))^K \approx B_i(\Sigma^K)$, resp. $(B_{n-i}(\Sigma_{(X)}^*))^K \approx B_i(\Sigma_{(K)}^{*K})$ gezeigt heben, so haben wir *das Pontrjagin'sche Theorem*:

Für die beiden topologischen Gruppensysteme Σ und $\Sigma_{(X)}^*$ gilt, dass die i-dimens. Bettigruppe eines jeden der beiden Systeme isomorph ist der K-Charaktergruppe der $(n-i)$-dimens. Bettigruppe des andern.

Als eine Verallgemeinerung der Relationen (I) und (II) geben wir

$$(\text{III}) \qquad (\Sigma_{(X)})^K \approx \Sigma_{(X^K)}^*,$$

für endliche Systeme $\sum$ und beliebige Gruppen X und K.

Definitionsgemäss ist $\sum x_\alpha a_\alpha^i$, $\alpha = 1, \cdots \rho_i$, $x_\alpha \subset X$ das allgemeine Element von $L_i(\Sigma_{(X)})$, d.h. diese Gruppe ist die direkte Summe der ρ_i Gruppen xa_α^i, deren jede X isomorph ist. Ist $f_{n-i} \subset (L_i(\Sigma_{(X)}))^K$, so gilt

$$(1) \qquad f_{n-i}(\sum x_\alpha a_\alpha^i) = \sum f_{n-i}(x_\alpha a_\alpha^i) = \sum f_\alpha(x_\alpha),$$

wo $f_\alpha \subset X^K$ durch

$$(2) \qquad f_\alpha(x) = f_{n-i}(xa_\alpha^i)$$

als K Charakter von X ($= xa_\alpha^i$) definiert ist. Ein Charakter. $f_{n-i} \subset (L_i(\Sigma_{(X)}))^K$ wird so in ein-eindeutiger Weise durch die ρ_i Charaktere $f_\alpha \subset X^K$, $\alpha = 1, \cdots \rho_i$, repräsentiert.

Wir ordnen zu $f_{n-i} \subset (L_i(\Sigma_{(X)}))^K = L_{n-i}(\Sigma_{(X)}^K)$ das Element $\sum f_\alpha a_\alpha^{*n-i} \subset L_{n-i}(\Sigma_{(X^K)}^*)$:

$$(3) \qquad f_{n-i} \leftrightarrow \sum f_\alpha a_\alpha^{*n-i}, \quad \text{wo} \quad f_\alpha(x) = f_{n-i}(xa_\alpha^i).$$

Diese Zuordnung ist, wie (3) anzeigt, ein-eindeutig und definiert eine Isomorphie $L_{n-i}(\Sigma_{(X)}^K) \approx L_{n-i}(\Sigma_{(X^K)}^*)$, deren Rand-Treue wir jetzt zeigen. In der Zuordnung (3) gilt (für $i+1$ statt i)

$$(4) \qquad R(f_{n-i}) \leftrightarrow \sum (Rf)_\beta a_\beta^{*n-i-1}, \quad \text{wo}$$

$$(5) \qquad \begin{aligned} (Rf)_\beta(x) &= (Rf_{n-i})(xa_\beta^{i+1}) = f_{n-i}(R(xa_\beta^{i+1})) = f_{n-i}\Big(\sum_\alpha \epsilon_{\beta\alpha}^{i+1} xa_\alpha^i\Big) \\ &= \sum_\alpha f_{n-i}(x\epsilon_{\beta\alpha}^{i+1} a_\alpha^i) = \sum_\alpha f_\alpha(x\epsilon_{\beta\alpha}^{i+1}) = \sum_\alpha \epsilon_{\beta\alpha}^{i+1} f_\alpha(x). \end{aligned}$$

Damit ist

$$(6) \qquad (Rf)_\beta = \sum_\alpha \epsilon_{\beta\alpha}^{i+1} f_\alpha = \sum \epsilon_{\alpha\beta}^{*n-i} f_\alpha,$$

und statt (4) gilt die Zuordnung

$$(7) \qquad R(f_{n-i}) \leftrightarrow \sum f_\alpha \epsilon^{*n-i}_{\alpha\beta} a^{*n-i-1}_\beta = R(\sum f_\alpha a^{*n-i}_\alpha);$$

damit ist (3) randtreu und (III) ist bewiesen.

Bemerkung: Sind X und K kontinuierliche Gruppen und definiert man wie in §7 die Topologie in den $L_i(\Sigma_{(X)})$, so gilt (III) auch bei Einschränkung auf stetige Charaktere.

Die Kombination von (III) mit dem Theorem des §8 führt zu der allgemeinen Relation

$$(8) \qquad B_i(\Sigma^*_{(Y^K)}) \approx (B_{n-i}(\Sigma_{(Y)}))^K,$$

für endliche Systeme Σ, *die gilt, wenn jeder K-Charakter* (resp. stetige *K.-Ch.*) *von* $_E Y$ *und von* EY (für jedes ganzzahl. E) *zu einem solchen von Y erweiterbar ist.*

(Im Fall der stetigen Charaktere sind Y und K *als kontinuierlich* und Y *zudem als kompakt* vorausgesetzt. Y^K bedeutet dann die Gruppe der *stetigen* K-Charaktere von Y.)

In (8) ist auch das Pontrjagin'sche Dualitätstheorem enthalten: Der erste Teil für $Y = G$, G = Gruppe der ganzen Zahlen, und K = Kreisgruppe, (dann ist $G^K \approx K$), der zweite Teil für $Y = K$ = Kreisgruppe, (dann ist $Y^K \approx G$, bei Einschränkung auf *stetige* K-Charaktere von Y).

§10. Zum Alexander'schen Dualitätstheorem

Σ sei ein topologisches Gruppensystem der Dimension $\leq n$ und M ein *Untersystem* von Σ. (Die i-dimens. Komplexgruppe $L_i(\mathrm{M})$ ist Untergruppe von $L_i(\Sigma)$, $i = 0, 1, \cdots$, und aus $a^i \subset L_i(\mathrm{M})$ folgt $R(a^i) \subset L_{i-1}(\mathrm{M})$).

Jeder K-Charakter von $L_i(\Sigma)$, d.h. Element f_{n-i} von $L_{n-i}(\Sigma^K) = (L_i(\Sigma))^K$ erzeugt einen K-Charakter $'f_{n-i}$ von $L_i(\mathrm{M})$, nämlich den mit dem er in den Elementen von $L_i(\mathrm{M})$ übereinstimmt:

$$'f_{n-i}(a^i) = f_{n-i}(a^i), \qquad a^i \subset L_i(\mathrm{M}).$$

Die so definierte Zuordnung der K-Charaktere

$$(1) \qquad f_{n-i} \to {}'f_{n-i}$$

stellt eine Homomorphie her: $L_{n-i}(\Sigma^K) \to L_{n-i}(\mathrm{M}^K)$, die randtreu ist und also eine Homomorphie der topologischen Gruppensysteme $\Sigma^K \to \mathrm{M}^K$ bedeutet.

In der Tat stimmen

$$R(f_{n-i})(a^{i+1}) = f_{n-i}(R(a^{i+1})) \quad \text{und} \quad R({}'f_{n-i})(a^{i+1}) = {}'f_{n-i}(R(a^{i+1}))$$

wegen $R(a^{i+1}) \subset L_i(\mathfrak{M})$ für $a^{i+1} \subset L_{i+1}(\mathfrak{M})$, d.h. in den Elementen von $L_{i+1}(\mathfrak{M})$ überein und entsprechen sich also in der (1) analogen Zuordnung: $R(f_{n-i}) \to R('f_{n-i})$.

Das Kernsystem der Homomorphie $\Sigma^K \to \mathfrak{M}^K$ *wird aus den* Kerngruppen der Homomorphien $L_{n-i}(\Sigma^K) \to L_{n-i}(\mathfrak{M}^K)$ gebildet. Seine $n - i$ dimens. Komplexgruppe $(L_{n-i}(\Sigma^K), L_i(\mathfrak{M}))$ wird aus den K-Charakteren von $L_i(\Sigma)$ gebildet, die in $L_i(\mathfrak{M})$ verschwinden, (§1).

Wir bezeichnen das *Kernsystem* entsprechend: $(\Sigma^K, \mathfrak{M})$.

Das Bildsystem der Homomorphie $\Sigma^K \to \mathfrak{M}^K$ wird aus den Bildgruppen der Homomorphien $L_{n-i}(\Sigma^K) \to L_{n-i}(\mathfrak{M}^K)$ gebildet. Seine $n - i$ dimens. Komplexgruppe $(L_{n-i}(\mathfrak{M}^K), L_i(\Sigma))$ wird aus den K-Charakteren von $L_i(\mathfrak{M})$ gebildet, die zu K-Charakteren von $L_i(\Sigma)$ erweiterbar sind, (§1).

Wir bezeichnen das Bildsystem entsprechend $(\mathfrak{M}^K, \Sigma)$.

In Formel:

$$L_{n-i}(\Sigma^K, \mathfrak{M}) = (L_{n-i}(\Sigma^K), L_i(\mathfrak{M})) \tag{2}$$

$$L_{n-i}(\mathfrak{M}^K, \Sigma) = (L_{n-i}(\mathfrak{M}^K), L_i(\Sigma)). \tag{2'}$$

Weiter gilt die Isomorphie der Gruppensysteme:

$$\Sigma^K - (\Sigma^K, \mathfrak{M}) \simeq (\mathfrak{M}^K, \Sigma), \tag{3}$$

welche Relation §1, (1) auf Gruppensysteme verallgemeinert.

Neben (3) besteht

$$(\Sigma^K, \mathfrak{M}) \simeq (\Sigma - \mathfrak{M})^K, \tag{4}$$

welche Relation §1, (3) auf Gruppensysteme erweitert.

Beweis von (4): Bezeichnet F_{n-i} resp. f_{n-i} ein Element von $L_{n-i}((\Sigma - \mathfrak{M})^K)$ resp. von $L_{n-i}(\Sigma^K, \mathfrak{M})$, so stellt die Zuordnung $f_{n-i} \leftrightarrow F_{n-i}$, wenn

$$f_{n-i}(a^i) = F_{n-i}(\{a^i\}_{L_i(\mathfrak{M})}), \qquad a^i \subset L_i(\Sigma), \tag{5}$$

ist, die Isomorphie her

$$L_{n-i}(\Sigma^K, \mathfrak{M}) \simeq L_{n-i}((\Sigma - \mathfrak{M})^K). \tag{6}$$

Da dann $R(f_{n-i}) = f_{n-i}(R(\ )) \subset L_{n-i-1}(\Sigma^K, \mathfrak{M})$ und $R(F_{n-i}) = F_{n-i}(R(\ )) \subset L_{n-i-1}((\Sigma - \mathfrak{M})^K)$ den (5) entsprechenden Relationen

$$f_{n-i}(R(a^{i+1})) = F_{n-i}(\{R(a^{i+1})\}_{L_i(\mathfrak{M})}) = F_{n-i}(R(\{a^{i+1}\}_{L_{i+1}(\mathfrak{M})})) \tag{7}$$

für $a^{i+1} \subset L_{i+1}(\Sigma)$ genügen, so ist die Zuordnung, die (6) herstellt, *randtreu*: d.h. (4) ist bewiesen.

Nach dieser Vorbereitung reduzieren wir die Allgemeinheit der Systeme Σ und M und der Gruppe K:

a) K sei die Kreisgruppe.

b) Σ sei ein freies topol. Gruppensystem mit endlicher Basis und M ein Untersystem von Σ, *dessen Basis ein Teilsystem ist der in* Σ *verwendeten Basis.*

c) Alle Bettigruppen $B_i(\Sigma)$ für $i = 1, \cdots n-1$, seien Nullgruppen.

Da alle Voraussetzungen für $B_i(\Sigma^K) \approx (B_{n-i}(\Sigma))^K$ erfüllt sind, sind auch die $B_i(\Sigma^K)$ für $i = 1, \cdots n-1$, Nullgruppen.

Damit gilt (Topol. Gruppensysteme)

$$\begin{aligned} &B_{i+1}(\Sigma - \mathrm{M}) \approx B_i(\mathrm{M}) \\ &B_{i+1}(\Sigma^K - (\Sigma^K, \mathrm{M})) \approx B_i(\Sigma^K, \mathrm{M}), \qquad i = 1, 2, \cdots n-2. \end{aligned} \tag{8}$$

Weiter ist

$$(\mathrm{M}^K, \Sigma) = \mathrm{M}^K \tag{9}$$

eine Folge der Voraussetzungen, da ja die Zerlegung in die *direkte* Summe

$$L_i(\Sigma) = L_i(\mathrm{M}) + L_i(N), \qquad i = 0, 1, \cdots \tag{10}$$

möglich ist. (Die $L_i(\mathrm{N})$ bilden *kein* topol. Untersystem von $\sum$).

Somit gilt (8), (3) und (9) für $i = 1, \cdots n-2$:

$$\begin{aligned} B_i(\Sigma^K, \mathrm{M}) \approx B_{i+1}(\Sigma^K - (\Sigma^K, \mathrm{M})) \approx B_{i+1}(\mathrm{M}^K, \Sigma) \\ \approx B_{i+1}(\mathrm{M}^K) \approx (B_{n-i-1}(\mathrm{M}))^K. \end{aligned} \tag{11}$$

Um dem Dualitätssatz seine endgültige Fassung zu geben, führen wir $\Sigma^*_{(K)}$ ein, wo (wie immer) Σ^* das zu Σ n-duale System bezeichnet.

Unter den Voraussetzungen (a), (b) und (c) bestehen nach §9 die Relationen

$$\Sigma^{\downarrow}_{(K)} \approx \Sigma^K \quad \text{und} \quad (\Sigma^*_{(K)})^K \approx \Sigma, \tag{12}$$

die letztere bei Einschränkung auf stetige K-Charaktere von $\Sigma^*_{(K)}$.

Bezeichnet $V_{(K)} \subset \Sigma^*_{(K)}$ das (Σ^K, M) in der Isomorphie $\Sigma^*_{(K)} \approx \Sigma^K$ entsprechende Untersystem von $\Sigma^*_{(K)}$, so folgt aus

$$V_{(K)} \approx (\Sigma^K, \mathrm{M}) \tag{13}$$

nach (4)

$$V_{(K)} \approx (\Sigma - \mathrm{M})^K. \tag{14}$$

Da infolge unserer Annahme $\Sigma - M$ *ein freies endliches System* ist, so haben wir (§9, (I))

$$(15) \qquad (\Sigma - M)^K \approx (\Sigma - M)^*_{(K)},$$

wo $(\Sigma - M)^*$ das n-duale System von $\Sigma - M$ ist. Daher gilt (14) und (15)

$$(16) \qquad V_{(K)} \approx (\Sigma - M)^*_{(K)}.$$

Da weiter (§9, (II))

$$(17) \qquad ((\Sigma - M)^*_{(K)})^K \approx \Sigma - M$$

bei Einschränkung auf stetige K-Charaktere besteht, so gilt §9 und (16), (17) und (8):

$$(18) \quad (B_i(V_{(K)}))^K \approx (B_i((\Sigma - M)^*_{(K)}))^K \approx B_{n-i}(\Sigma - M) \approx B_{n-i-1}(M), \quad i = 1, \cdots n-2,$$

bei Einschränkung auf stetige K-Charaktere.

Zusammenfassend ist damit in

$$(19) \qquad \begin{cases} B_i(V_{(K)}) \approx (B_{n-i-1}(M))^K, \\ (B_i(V_{(K)}))^K \approx B_{n-i-1}(M), \qquad i = 1, \cdots n-2, \end{cases}$$

das Alexander'sche Dualitätstheorem in der Pontrjagin'sche Fassung ausgesprochen.

Bemerkung: *In der Isomorphie* (16) *haben entsprechende Elemente gleiche Darstellungszahlen.*

In der Tat besteht

$$(20) \qquad (\Sigma^K, M) \approx (\Sigma - M)^K$$

infolge der Zuordnungen

$$(21) \qquad \begin{gathered} f_i(a^{n-i}) = F_i(\{a^{n-i}\}_{L_{n-i}(M)}), \\ f_i \subset L_i(\Sigma^K, M), \qquad F_i \subset L_i((\Sigma - M)^K). \end{gathered}$$

Nun induzieren (§9) die K-Charaktere f_i resp. F_i infolge der Isomorphien $\Sigma^K \approx \Sigma^*_{(K)}$, resp. $(\Sigma - M)^K \approx (\Sigma - M)^*_{(K)}$, i-dimens. Komplexe in $\Sigma^*_{(K)}$ resp. $(\Sigma - M)^*_{(K)}$ mit den respektiven Darstellungszahlen

$$(22) \qquad f_i(a^{n-i}_\alpha) = y_\alpha, \qquad F_i(\{a^{n-i}_\alpha\}_{L_{n-i}(M)}) = y_\alpha,$$

die wegen (21) übereinstimmen.

(Dabei verschwindet f_i für ein Basiselement a_α^i von M,—für welches F_i undefiniert ist.)

Von der in der Bemerkung ausgesprochenen Tatsache wurde in (18) bereits Gebrauch gemacht.

§11. Ergänzungen

Bei der Behandlung spezieller Systeme Σ, für die bezüglich der gegebenen Gruppe K

$$(1) \qquad (B_i(\Sigma))^K \approx B_{n-i}(\Sigma^K)$$

gilt, haben wir uns auf Systeme (der Dimens. $\leq n$) mit endlicher Basis beschränkt.

Die Benutzung der Resulte der Arbeit "Theorie of Topological Groups" von Pontrjagin (Annals of Math. Vol. 35, 1934) gestattet die Ausdehnung gewisser unserer Ergebnisse auf Systeme Σ *mit abzählbarer Basis.*

Es *gilt* z. B. (1) *für ein solches freies System, wenn K die Kreisgruppe ist.* ($K^i \subset L_i(\Sigma)$) hat die Darstellung: $K^i = \sum x_\alpha a_\alpha^i$, $\alpha = 1, 2, \cdots \infty$, wo aber nur endlich viele der ganzzahligen x_α von Null verschieden sind.)

In der Tat enthält das Lemma 1 der erwähnten Arbeit die Tatsache der Erweiterbarkeit jedes K-Charakters jeder Untergruppe von $L_i(\Sigma)$ zu einem solchen von $L_i(\Sigma)$. Aus dem Hauptergebnis der Pontrjagin'schen Arbeit folgt dann weiter

$$(1') \qquad (B_i(\Sigma^K))^K \approx B_{n-i}(\Sigma),$$

dies bei Einschrankung auf "stetige" K-Charaktere von $B_i(\Sigma^K)$.

Dabei wird in $L_i(\Sigma^K)$ (und damit in $B_i(\Sigma^K)$ eine Limes-Vorschrift eingeführt durch:

$$(\alpha) \quad \lim_{m=\infty} f_{mi} = f_i \qquad \text{wenn} \qquad \lim_{m=\infty} f_{mi}(K^{n-i}) = f_i(K^{n-i}) \qquad \text{für jedes}$$

beliebige Element K^{n-i} von $L_{n-i}(\Sigma)$ gilt.

(Da offenbar aus $\lim f_{mi} = f_i$ die Relation $\lim (Rf_{mi}) = Rf_i$ folgt, so ist in bezug auf die in den $L_i(\Sigma^K)$ durch (α) eingeführte Topologie der Rand- Operator $R(\ )$ stetig.)

Wieder ist Σ^K *algebraisch* und *topologisch* isomorph $\Sigma^*_{(K)}$, wo Σ^* das Σ n-duale System ist, wenn man in $\Sigma^*_{(K)}$ eine Topologie einführt durch:

$$(\beta) \quad \lim_{m=\infty} K_m^{*i} = K^{*i}, \qquad \text{wenn} \qquad \lim_{m=\infty} y_{m\alpha} = y_\alpha, \qquad \alpha = 1, 2, \cdots \infty$$

gilt für die Darstellungszahlen der $K_m^{*i} = \sum y_{ma} a_a^{*i}$, resp. des K, $= \sum y_a a_a^{*i}$.

Damit gilt das Hauptergebnis des §9 *auch für Systeme* $\sum$ *mit abzählbarer Basis* (K = Kreisgruppe).

(Σ ist dabei auf endliche Komplexe beschränkt, wogegen für die Komplexe der $L_i(\Sigma^*_{(K)})$ keine solche Einschränkung besteht. Die $L_i(\Sigma^*_{(K)})$ sind offenbar kompakt.)

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

A NOTE ON TRANSCENDENTAL NUMBERS

By Brockway McMillan

Introduction. Liouville in 1851 proved the transcendence of numbers of the form $\sum_{n=0}^{\infty} a_n/l^{n!}$ where the a_n are bounded integers and l is an integer ≥ 2. Since that time many other infinite series have been considered which lead to transcendental numbers. Illustrative of these is that considered by Kempner:[1]

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{b^{n!}} x^n$$

which is transcendental for every rational x, when a, b are integers, and the a_n are (say) bounded integers of which only a finite number vanish. All these series have terms whose denominators increase very rapidly with n. The transcendence[2] of $e^x = \sum x^n/n!$ for every algebraic $x \neq 0$, exhibits a series of terms with much less rapidly increasing denominators which leads to a transcendental number. Mahler[3] has shown that numbers of the form $\sum f(n)/q^{\lambda_n}$ are transcendental where $f(n)$ is a polynomial in n with integral coefficients, q is any integer > 1, and $\lambda_{n+1} - \lambda_n$ equals the number of digits in $f(n)$.

In this note we exhibit two classes of transcendental numbers defined by infinite series: $\sum_{-\infty}^{\infty} f(n)$ where $f(z)$ is a rational function of z with certain restrictions on the nature of its poles; and $\sum_{0}^{\infty} a_n z^n/n!$ for all algebraic $z \neq 0$, where the chief restriction on the a_n is that they be algebraic numbers satisfying some recursion relationship: $\sum_{n} c_n a_{n+p} = 0$, all p. The methods we use are not those of transcendence theory; we sum the series in question explicitly in terms of the known transcendental numbers e and π.

We define an algebraic polynomial (in z) to be a polynomial (in z) all of whose coefficients are algebraic numbers.

[1] Kempner, A. J., Trans. A. M. S. *17* 476 (1916). He refers here to others.
[2] Lindemann, E., Mathematische Annalen *20* 213 (1882).
[3] Mahler, K., Amst. Proc. *40* 420 (1937).

Theorem I. Let $f(z) \equiv P(z)/Q(z)$ be a rational function of z where $P(z)$ is an algebraic polynomial in z of degree p, and $Q(z)$ an algebraic polynomial of degree $q \geq p + 2$. Further, let the zeros of $Q(z)$ be at the points $z = \zeta_k$, of orders ν_k respectively $(k = 1, 2, \cdots K)$, where the ζ_k are real rational numbers, not integers.

Under these hypotheses, the sum

$$\sum_{n=-\infty}^{\infty} f(n)$$

if it does not vanish, represents a transcendental number.

We sum the series by a well known method. Consider the function $\pi \cot \pi z\, f(z)$. This has poles at the integer points $z = n$ with residues $f(n)$, and poles at the points $z = \zeta_k$ with (say) residues b_k. Let C_N be the square contour with corners at $(N + \frac{1}{2})(\pm 1 \pm i)$, N an integer. For $N \geq \max |\zeta_k|$

$$\frac{1}{2\pi i} \int_{C_N} \pi \cot \pi z\, f(z) = \sum_{n=-N}^{N} f(n) + \sum_{k=1}^{K} b_k$$

It is a simple matter to show that the contour integral can be made arbitrarily small in absolute value by choosing N sufficiently large, from which it follows that

$$\sum_{n=-\infty}^{\infty} f(n) = -\sum_{k=1}^{K} b_k$$

Now b_k is the residue of $\pi \cot \pi z\, f(z)$ at $z = \zeta_k$. Since this point is a pole of order ν_k, we may write

$$\pi \cot \pi z\, f(z) = \frac{\pi \cot \pi z\, g_k(z)}{(z - \zeta_k)^{\nu_k}}$$

where $g_k(z)$ is regular in some neighborhood of $z = \zeta_k$. Since the ζ_k are finite in number they are uniformly bounded away from the integer points, and $\pi \cot \pi z\, g_k(z)$ is therefore regular within and upon some contour γ_k enclosing only the pole at $z = \zeta_k$. Then

$$b_k = \frac{1}{2\pi i} \int_{\gamma_k} \pi \cot \pi z\, f(z)\, dz = \frac{1}{2\pi i} \int_{\gamma_k} \frac{\pi \cot \pi z\, g_k(z)}{(z - \zeta_k)^{\nu_k}}\, dz$$

Or, applying Cauchy's formula

$$b_k = \frac{1}{(\nu_k - 1)!} \frac{d^{\nu_k - 1}}{dz^{\nu_k - 1}} \{\pi \cot \pi z\, g_k(z)\}_{z=\zeta_k}$$

Using Leibniz's Rule

$$b_k = \frac{\pi}{(\nu_k - 1)!} \sum_{p=0}^{\nu_k - 1} C_{\nu_k - 1}^{p} \left[\frac{d^p}{dz^p} \cot \pi z\right]\left[\frac{d^{\nu_k - p - 1}}{dz^{\nu_k - p - 1}} g_k(z)\right]_{z=\zeta_k}$$

Now $g_k(z) = (z - \zeta_k)^{\nu_k} f(z)$ is a rational function of z with algebraic coefficients. The same, therefore, is true of all of its derivatives, which, accordingly, take on algebraic values for the algebraic argument ζ_k. Also, $\cot \pi z = (e^{2i\pi z} + 1)(e^{2i\pi z} - 1)^{-1}$ is a rational function of $e^{2i\pi z}$, with algebraic coefficients. $\dfrac{d^p}{dz^p} \cot \pi z$ therefore is π^p times a similar function.

Since $e^{2i\pi z}$ takes on algebraic values for each $z = \zeta_k$ (herein lies the necessity that the ζ_k be real and rational), it follows that each non zero b_k is an algebraic polynomial in π, in which the lowest power of π is at least the first. Since by hypothesis not all the b_k vanish, we conclude that their sum is an algebraic polynomial in π in which at least one non zero power of π appears. This sum, which is $-\sum_{-\infty}^{\infty} f(n)$, is therefore a transcendental number.

Theorem II. Given the set of algebraic numbers $a_n (n = 0, 1, 2, \cdots)$ of which an infinite number are non-vanishing, and such that

$$\limsup_{n \to \infty} |a_n|^{1/n} \leq \rho \tag{1}$$

Let there be a set of constants $c_n (n = 0, 1, 2, \cdots)$, at least two of which are non-vanishing, such that the function

$$g(t) \equiv \sum_{n=0}^{\infty} c_n t^n \tag{2}$$

is analytic for $|t| < R$ where $R > \rho$, and has zeros for $|t| \leq \rho$ only at the points $t = \zeta_k$ $(k = 1, 2, \cdots N)$ respectively of multiplicities ν_k, where the ζ_k are algebraic numbers. If now

$$\sum_{n=0}^{\infty} c_n a_{n+p} = 0 \tag{3}$$

for all $p = 0, 1, 2, \cdots$, then the function

$$F(z) \equiv \sum_{n=0}^{\infty} \frac{a_n z^n}{n!} \tag{4}$$

takes on a transcendental value for every algebraic $z \neq 0$.

It is known that, under the above hypotheses, the function $F(z)$ can be expressed in the form

$$F(z) = \sum_{k=1}^{N} P_k(z)e^{\zeta_k z} \tag{5}$$

where $P_k(z)$ is a polynomial in z of degree $\nu_k - 1$. If only a finite number of the c's are non-vanishing—that is, if the recursion relationship (3) is of finite order—(5) follows immediately from the fact that $F(z)$ satisfies an ordinary linear differential equation with constant coefficients: $\sum c_n F^{(n)}(z) = 0$. In the somewhat more general case here treated, (5) may be deduced by the methods used in handling linear differential equations of infinite order.[4]

To prove our theorem we need only show that the polynomials $P_k(z)$ are algebraic polynomials, for if they have algebraic coefficients, with every algebraic $z \neq 0$, $F(z)$ becomes a linear combination,[5] with algebraic coefficients, of algebraic powers of e. Since at least one non zero power of e appears ($F(z)$ is not a polynomial in z, by hypothesis), by Lindemann's theorem[6] $F(z)$ is then a transcendental number.

Let the coefficients of $P_k(z)$ be X_{kp}, so that (5) becomes

$$F(z) = \sum_{k=1}^{N} \sum_{p=0}^{\nu_k - 1} X_{kp} z^p e^{\zeta_k z} \tag{6}$$

Let $\sum_{k=1}^{N} \nu_k = M + 1$. Consider the $M + 1$ equations (from (4))

$$\frac{d^m}{dz^m} F(z) \Big|_{z=0} \equiv F^{(m)}(0) = a_m \qquad m = m_0, m_0 + 1, \cdots m_0 + M$$

where m_0 is so chosen that at least one of $a_{m_0} \cdots a_{m_0+M}$ is not zero. These equations in terms of the X_{kp} are easily obtained from (6):

$$\sum_{k=1}^{N} \sum_{p=0}^{\nu_k - 1} m(m-1)\cdots(m-p+1)\zeta_k^{m-p} X_{kp} = a_m \tag{7}$$

$$m = m_0, m_0 + 1, \cdots m_0 + M$$

[4] E.g., O. Perron, Mathematische Annalen *84* 31 (1921).

[5] We use throughout the fact that the algebraic numbers form a field.

[6] Lindemann, E., loc. cit. Strictly speaking, Lindemann's theorem states that no linear combination with algebraic coefficients, of distinct algebraic powers of e can vanish. It follows immediately that such an "exponential form" cannot represent an algebraic number.

These are $M + 1$ non homogeneous linear equations in the $M + 1$ quantities X_{kp}. The coefficients of the X_{kp} are clearly algebraic numbers, as are the a_m. If equations (7) can be solved for the X_{kp} in terms of the a_m and the coefficients, the X_{kp} are algebraic numbers and our result is established.

The equations (7) may be written

$$(8) \qquad \sum_{k=1}^{N}\left[\sum_{p=0}^{\nu_k-1}\left\{\frac{d^p}{dz^p}\,z^m\right\}_{z=\zeta_k} X_{kp}\right] = a_m \qquad m = m_0, m_0+1, \cdots m_0+M$$

The determinant of coefficients of this system can now be written down in the following way: a column of the powers of ζ_1, $-\zeta_1^{m_0}$ to $\zeta_1^{m_0+M}$, $-$ is written down. If $\nu_1 = 1$, the next column is formed by writing the same powers of ζ_2. If $\nu_1 \neq 1$, the elements of the second column are the first derivatives (with respect to ζ_1) of the corresponding elements of the first column. This process is repeated, each column being the derivative of the one on its left, until ν_1 columns have been filled. Then a column of the powers of ζ_2, $-\zeta_2^{m_0}$ to $\zeta_2^{m_0+M}$, $-$ follows with its $\nu_2 - 1$ derivatives, and so on. We now show that this determinant does not vanish.

If the determinant did vanish, it would follow that the rows were linearly dependent. Suppose this to be the case. Then there exists a set of constants A_s such that the ν_k equations

$$\sum_{s=0}^{M} A_s x^{m_0+s} = 0$$

$$\sum_{s=0}^{M} (m_0+s) A_s x^{m_0+s-1} = 0$$

$$\vdots$$

$$\sum_{s=0}^{M} (m_0+s)(m_0+s-1)\cdots(m_0+s-\nu_k+2) A_s x^{m_0+s-\nu_k+1} = 0$$

are all satisfied for $x = \zeta_k$. That is to say, the equation

$$(9) \qquad \sum_{s=0}^{M} A_s x^{m_0+s} = 0$$

admits each root $x = \zeta_k$ with multiplicity at least ν_k. But (9) admits the root $x = 0$ with multiplicity m_0, so the degree of (9) must be at least

$$m_0 + \sum_{k=1}^{N} \nu_k = m_0 + M + 1$$

But (9) is clearly of degree not greater than $m_0 + M$. Consequently, the determinant of coefficients in equations (7) cannot vanish, the quantities of X_{kp} are algebraic, and the polynomials $P_k(z)$ are algebraic polynomials.

We append a few remarks on theorem II, without proofs:

1. In assuming that the recursion relationship (3) is of possibly infinite order, we have not actually included a larger class of functions $F(z)$ than would have been included had we admitted only recursion relationships of finite order. Every set of a_n satisfying our hypotheses satisfies a finite recursion relationship

$$\sum_{m=0}^{M+1} c_m^* a_{m+p} = 0 \qquad p = 0, 1, 2, \cdots \tag{10}$$

where the c_m^* are such that the equation

$$\sum_{m=0}^{M+1} c_m^* x^m = 0 \tag{11}$$

has roots $x = \zeta_k$ each with multiplicity ν_k.

2. We have proved essentially that any function $F(z)$ satisfying a linear differential equation with constant algebraic[7] coefficients

$$\sum_{m=0}^{M+1} c_m^* F^{(m)}(z) = 0$$

and meeting the algebraic "initial conditions"

$$F^{(m)}(0) = a_m \qquad m = 0, 1, \cdots M$$

takes on a transcendental value for each algebraic argument $z \neq 0$.

3. Any $\bar{F}(z)$ whose coefficients $\bar{a}_m$ are algebraic numbers satisfying our original hypotheses from some $\bar{a}_{m_0}$ on differs from $F(z)$ only by an algebraic polynomial in z of degree at most $m_0 - 1$, and is therefore also transcendental for every algebraic $z \neq 0$.

Finally, we point out that the two classes of transcendental numbers which we have considered are both denumerable.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

[7] The c_m^* need not be algebraic, so long as (11) has only algebraic roots. $F(z)$, by our original hypotheses, satisfied a differential equation of infinite order.

A COMPLETE SET OF ASYMPTOTIC FORMULAS FOR THE WHITTAKER FUNCTION AND THE LAGUERRE POLYNOMIALS

By W. C. Taylor[1]

In the following pages there are obtained formulas asymptotically describing the Whittaker function $W_{k,m}(z)$ of the complex variables k, m, and z as $k \to \infty$. The results are complete in the sense that no restriction is imposed on the behavior of z as $k \to \infty$. The variable m remains bounded. The results are given in formulas (26a, b, c, d), (27), (32), (33), (34), and are specialized and simplified to give descriptions of the Laguerre polynomials in (35) to (40).

Whittaker's confluent hypergeometric differential equation,[2]

$$\frac{d^2u}{dz^2} + \left\{-\frac{1}{4} + \frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right\}u = 0, \tag{1}$$

has solutions given by

$$M_{k,\pm m}(z) = z^{\frac{1}{2}\pm m}e^{-\frac{z}{2}}\left\{1 + \frac{\frac{1}{2} \pm m - k}{1!(\pm 2m + 1)}z + \frac{(\frac{1}{2} + m - k)(\frac{3}{2} + m - k)}{2!(\pm 2m + 1)(\pm 2m + 2)}z^2 + \cdots\right\}, \tag{2}$$

and

$$W_{k,m}(z) = \frac{\Gamma(-2m)}{\Gamma(\frac{1}{2} - m - k)}M_{k,m}(z) + \frac{\Gamma(2m)}{\Gamma(\frac{1}{2} + m - k)}M_{k,-m}(z). \tag{3}$$

Since (1) is unchanged by a simultaneous replacement of z and k by $-z$ and $-k$, $W_{-k,m}(-z)$ and $M_{-k,\pm m}(-z)$ are also solutions.

Whittaker obtained an asymptotic expansion (of which only the first term is reproduced here),

$$W_{k,m}(z) = e^{-\frac{z}{2}}z^k[1 + O(z^{-1})], \tag{4}$$

valid as $z \to \infty$ with $|\arg z| < \pi - \epsilon < \pi$, k and m remaining bounded. The asymptotic evaluation in the case when $k \to \infty$ has presented greater

[1] University of Wisconsin thesis, 1936.

[2] Whittaker, E. T. and Watson, G. N.: *Modern Analysis*, 4th edition, chap. XVI. Gives equations (1) to (4).

difficulty. It has been attacked by application of the saddle point method to contour integrals by Perron,[3] who obtained an asymptotic expansion (as distinct from a *formula* of one term plus remainder such as (4)) for the case in which z remains bounded; by Frenkel,[4] who treated the case $m = \pm\frac{1}{2}$; and by Erdelyi,[5] who gives a special case of formula (33). Schwid[6] has used the same methods as the present paper in studying the case $m = \pm\frac{1}{4}$. Schwid's work and results are almost identical with those of the present paper through sections 1 to 4, since m is not significant in the formulas obtained in these sections for z relatively large.

The (generalized) Laguerre polynomials defined by

$$L_n^{(\alpha)}(z) = \frac{1}{n!} z^{-\alpha} e^{z} \frac{d^n}{dz^n} (z^{n+\alpha} e^{-z})$$

are included in the Whittaker function as the set of polynomials represented by the brace in (2):

$$L_n^{(\alpha)}(z) = \frac{\Gamma(n+\alpha+1)}{n!\Gamma(\alpha+1)} z^{-\frac{\alpha+1}{2}} e^{\frac{z}{2}} M_{n+\frac{\alpha+1}{2},\frac{\alpha}{2}}(z) = \frac{(-1)^n}{n!} z^{-\frac{\alpha+1}{2}} e^{\frac{z}{2}} W_{n+\frac{\alpha+1}{2},\frac{\alpha}{2}}(z). \tag{5}$$

The need of asymptotic evaluations of these polynomials in the study of convergence and summability problems has motivated a number of investigations.[7] The formulas to be obtained here contain the results (concerning asymptotic descriptions) of the literature cited in the footnote with several improvements of the error terms and with new results

[3] Journal für r. u. a. Math. 151 (1921), pp. 63–78.

[4] Frenkel, M.: *Die Asymptotischen Losungen der in der Theorie der radioaktiven Emission auftretenden Differentialgleichung.* Zeitschrift für Physik, 95 (1935), p. 599.

[5] Erdelyi, Artur: *Über eine Integraldarstellung der $M_{k,m}$ Funktionen und ihre asymptotische Darstellung für grosse Werte von $Re(k)$.* Math. Ann. 113 (1936), p. 357.

[6] Schwid, Nathan: *On the asymptotic forms of the Hermite and Weber functions,* Trans. Am. Math. Soc., 37 (1935), p. 339.

[7] Fejer, L.: Comptes Rendus 147 (1908), p. 1040. Szegö, G.: Mathematische Zeitschrift, 25 (1926), p. 87. Uspensky, J. V.: Annals of Mathematics, 28 (1927). Wright, E. M.: Journal of the London Mathematical Society, 7 (1932), p. 256. Hlawka, E.: Monatshefte für Mathematik u. Physik., 42 (1935), p. 275. Kogbetliantz, E.: This Journal, XIV (1935), p. 51. Ottaviani, G.: Ann. Scuola norm. super. Pisa II s. 7 (1938), p. 1. Literature concerned with the asymptotic character of the Hermite polynomials has been omitted.

for the range, z real, $z > kn$, $k < 4$. There are two exceptions, earlier results not contained in this paper: Uspensky gives an added explicit term for the formula (38) for the case $z = O(1)$, leaving a remainder $O(n^{-1})$. Wright gives a complete expansion, an extension of (36) or (37) for $z = O(1)$ complex, with the remark that terms after the first "can be calculated with sufficient labor."

The entire set of formulas is brought forth here by two straightforward applications of a single explicit method. This method, or theorem, given by Langer,[8] is stated in section 1, taken from Langer [2]. In sections 2, 3, 4, the application is made to the Whittaker function to obtain an asymptotic description[9] valid as $k \to \infty$, $|z| > N|k|^{-1+\epsilon}$. In sections 5 and 6 the methods of Langer [3] are applied to obtain an asymptotic description when $z = O(k^{1-\epsilon})$. Finally, in section 7, asymptotic formulas for the Laguerre polynomials $L_n^{(\alpha)}(x)$, with x real and positive, are set down as they may be obtained by simplification of the preceding, more general formulas.

1. **The general theorem.** For convenience of reference, those of the results of Langer [2] which are to be used here will be summarized informally as a theorem. In addition the theorem to be stated broadens those results by permitting the region (called R_s) in which the investigation is carried out, to expand as $|\rho|$ increases. The orders of the remainder terms are correspondingly changed. The discussion given by Langer suffices for this more inclusive statement with only small and obvious changes.

The equation

$$\frac{d^2u}{ds^2} + \{\rho^2\varphi^2(s) + \chi(s)\}u = 0 \tag{6}$$

is under consideration in a simply connected region $R_s(\rho)$ of the plane of the complex variable s. This region is to contain the point $s = a$, at which $\varphi^2(s)$ vanishes. The discussion is concerned only with sufficiently large values of the complex variable ρ. In its dependence on ρ, the region $R_s(\rho)$ is to be permitted to swell as $|\rho|$ increases, subject only to the restriction of (V) below.

[8] Langer, R. E.: Trans. Am. Math. Soc., vol. 33 (1931), p. 23. Also vol. 34 (1932), p. 447 and vol. 37 (1935), p. 397. The second and the third papers each include, extend and simplify the results of their predecessors. They will be designated by Langer [2] and Langer [3] respectively.

[9] ϵ and N will each be used to denote "some positive constant," ϵ small and N sufficiently large. They do not maintain their identities from one expression to the next.

The following hypotheses are to be fulfilled in $R_s(\rho)$:

(I) *$\varphi^2(s)$ is an analytic function of the form*

$$\varphi^2(s) = (s - a)^\nu \varphi_1^2(s), \qquad \nu \geq 0$$

where, in a subregion defined by $| s - a | < \eta$, η a small positive number, $\varphi_1^2(s)$ is single-valued, bounded, and bounded from zero.

(II) *$\chi(s)$ is analytic and, when $| s - a | < \eta$, is bounded.*

New variables are introduced:

$$\Phi(s) = \int_a^s \varphi(s)\, ds \tag{7}$$

and

$$\xi = \rho\Phi(s)$$

(III) *$\xi \to \infty$ as $\rho \to \infty$ if s remains in the portion $| s - a | > \eta$ of $R_s(\rho)$.*

The relations (7) define a mapping of the region (of the Riemann surface) $R_s(\rho)$ on the two corresponding regions $R_\Phi(\rho)$ and $R_\xi(\rho)$. The map R_ξ is evidently obtained from R_Φ by an expansion and rotation about the origin.

(IV) *For each value of ρ, there is, in the part of $R_\xi(\rho)$ in any quadrant*

$$p\,\frac{\pi}{2} \leq \arg \xi \leq (p + 1)\,\frac{\pi}{2}, \qquad p \text{ an integer},$$

a point $Q(\rho)$ at which $| I(\xi) |$ is maximum (possibly infinite),[10] *and a set of ordinary curves Γ, having the properties*: a) *$I(\xi)$ is monotone (non-increasing or non-decreasing) along each curve of Γ.* b) *Each point of R_ξ in this quadrant is joined to Q by a curve of Γ.* c) *Each such point is joined to the origin by a curve of Γ.*

There is defined further:

$$\mu = \frac{1}{\nu + 2},$$

and

$$\theta(s) = \chi(s) - \frac{1}{4\Omega^2}\{(\Omega^2)'' - 3(\Omega')^2 + 4\mu^2\}, \tag{8}$$

where

$$\Omega(s) = \frac{\Phi(s)}{\varphi(s)}.$$

[10] $R(w)$ and $I(w)$ are to denote, respectively, the real and imaginary parts of w.

(V) *The curves* Γ *may further be chosen so that*

$$\int \left| \frac{\theta(s)}{\varphi(s)}\, ds \right| = O(\rho^{\alpha}), \qquad 0 \leq \alpha < 1, \quad (9)$$

uniformly, where the integral is taken along the portion on which $|s - a| > \eta$ *of any one of those curves.*

An equation (6) which satisfies the preceding hypotheses is approximated by the equation

$$\frac{d^2y}{ds^2} + \{\rho^2\varphi^2(s) + \chi(s) - \theta(s)\}y = 0,$$

with solutions

$$y = \Phi^{\frac{1}{2}}(s)\varphi^{-\frac{1}{2}}(s)C_{\mu}(\xi)$$

C_{μ} being any Bessel function of order μ. And on the basis of this approximation there can be deduced the existence in $R_s(\rho)$ of independent pairs of solutions for (6) which have the forms:[11]

$$\begin{aligned} u_{k,j}(s, \rho) &= \rho^{\mu-\frac{1}{2}}\varphi^{-\frac{1}{2}}e^{\pm i\xi}[1 + O(\rho^{-\tau}) + O(\xi^{-1})], \\ k &= 0, \pm 1, \pm 2, \cdots \\ j &= 1 \text{ or } 2 \end{aligned} \quad (10)$$

when ξ remains in the sector $\Xi^{(k)}$ defined by $|\arg \xi - k\pi| < \pi - \epsilon$, $\epsilon > 0$. Here r is the smaller of the two numbers

$$1 - \alpha \quad \text{and} \quad \begin{cases} 1 & \text{if } \mu > \frac{1}{2} \\ 1 - \delta,\ \delta > 0 \text{ but arbitrarily small,} & \text{if } \mu = \frac{1}{2}. \\ 4\mu, & \text{if } \mu < \frac{1}{2}. \end{cases}$$

These solutions $u_{k,j}$ are not necessarily all analytic functions of ρ. In general

$$\begin{aligned} u_{k,j}(s, \rho) = \rho^{\mu-\frac{1}{2}}\varphi^{-\frac{1}{2}}[&A_{j,1}^{k,\Lambda}e^{i\xi}\{1 + O(\rho^{-\tau}) + O(\xi^{-1})\} \\ &+ A_{j,2}^{k,\Lambda}e^{-i\xi}\{1 + O(\rho^{-\tau}) + O(\xi^{-1})\}] \end{aligned} \quad (11)$$

[11] The convention is adopted here that in any relation containing the letter j and a double sign, as in (10), the upper sign is to be taken when $j = 1$ and the lower when $j = 2$.

The constants implied in the O terms are independent of all the variables involved, depending only on the choices of arbitrary constants which have been made. This will be true of all such terms throughout this paper.

when ξ remains in the sector $\Xi^{(h)}$, where

$$A_{j,m}^{h,h} = \frac{\pm\pi}{i \sin \mu\pi} [c_{2,3-j}^{(h)} c_{1,m}^{(h)} - c_{1,3-j}^{(h)} c_{2,m}^{(h)}],$$
$$c_{i,j}^{(h)} = (2\pi)^{-\frac{1}{2}} e^{[2h+(-1)^{h+i}](\frac{1}{2}\mp\mu)\frac{\pi i}{2}}. \tag{12}$$

And if ξ remains bounded as ρ becomes infinite,

$$u_{h,j}(s, \rho) = \frac{\pm\pi}{i \sin \mu\pi} \rho^{\mu} \Phi^{\frac{1}{2}} \varphi^{-\frac{1}{2}} \cdot \{c_{2,3-j}^{(h)} J_{-\mu}(\xi) - c_{1,3-j}^{(h)} J_{\mu}(\xi)\} + O(\rho^{-1}) \tag{13}$$

where $J_{\pm\mu}$ denotes the Bessel function of the first kind.

2. **Application of the theorem.** The theorem of the preceding section will now be applied to the determination of asymptotic formulas for the Whittaker function.

In terms of the variable

$$s = \frac{z}{4k} \tag{14}$$

(1) becomes

$$\frac{d^2u}{ds^2} + \left\{4k^2 \frac{1-s}{s} + \frac{\frac{1}{4} - m^2}{s^2}\right\} u = 0 \tag{15}$$

This is of the form (6) with

$$\rho = 2ik, \qquad \varphi^2(s) = \frac{s-1}{s}, \qquad \chi(s) = \frac{m^2 - \frac{1}{4}}{s^2} \tag{16}$$

The zero of $\varphi^2(s)$ at $s = 1$ is of first order:

$$a = 1, \qquad \nu = 1, \qquad \mu = \tfrac{1}{2}.$$

Equation (7) becomes

$$\Phi(s) = \int_1^s \sqrt{\frac{s-1}{s}}\, ds$$
$$= \sqrt{s(s-1)} - \log(\sqrt{s} + \sqrt{s-1}). \tag{17}$$

This function Φ has branch points at $s = 0, 1, \infty$. It will be shown that the five hypotheses are satisfied in a region $R_s(\rho)$, formed by removing a neighborhood (depending on ρ) of $s = 0$ from the portion defined by

$$-\pi \leq \arg s \leq \pi \tag{18}$$

of the Riemann surface of the variable s for the function Φ. Since $\arg s = \arg z - \arg k$ and the functions (2) and (3) under consideration are single valued in k, the restriction (18) does not impoverish the results.

3. **The hypotheses.** Hypotheses (I) and (II) are satisfied in this region. It will be shown that Φ is bounded from zero in any portion $|s-1| > \eta > 0$ of R_s, and from this the fulfillment of (III) is evident. Now if the path of integration in (17) be chosen as a straight line, $\arg(s-1) = \arg ds$ is constant. The variation of $\arg s$ can be at most π along this path; that of $\arg d\Phi$, at most $\frac{\pi}{2}$. The conclusion that Φ is bounded from zero follows.

Along the boundary fixed by (18), the negative real axis of the s plane, the function $\varphi = \frac{d\Phi}{ds}$ is real; consequently this boundary is mapped into a line $I(\Phi) = \text{constant}$. Further, at $s = 0$ the mapping is non-conformal, angles being reduced to one-half. The angle of size 2π of the 'corner' of the region R_s at $s = 0$ is mapped into a straight angle in the Φ plane. The whole boundary, then, maps into a complete line parallel to the real axis, with an exception necessitated by the exclusion of the neighborhood of $s = 0$. The map is shown in the figure opposite.

The map in the ξ plane is obtained now by a rotation and expansion about $\Phi = 0$ of the configuration R_Φ. The point Q of (IV) may be chosen, in each quadrant, for each value of $\arg \rho$, either on one of the bounding lines or at infinity. If the excluded neighborhood of $s = 0$ be chosen of suitable shape, depending on $\arg \rho$, the curves Γ may be taken as straight lines except possibly for small kinks about the excluded neighborhood.

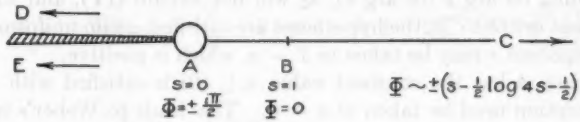
It remains to show that (V) is fulfilled by this choice. The relation (9) is seen directly to be satisfied with $\alpha = 0$, excepting at most the immediate neighborhoods of $s = 0$ and $s = \infty$. About $s = \infty$, the following convergent expansions may be deduced from (16), (17), and (8):

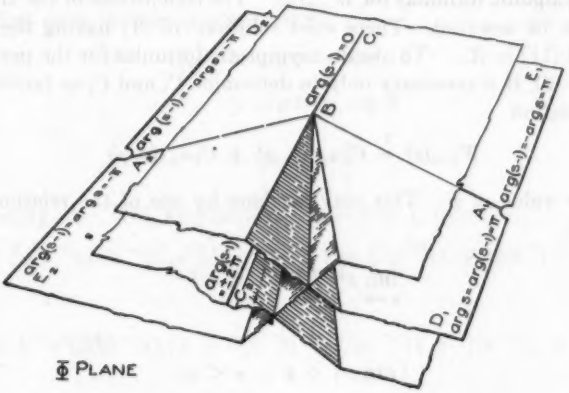
$$\begin{aligned} \varphi(s) &= \pm\left(1 - \frac{1}{2s} - \frac{1}{8s^2} + \cdots\right). \\ \Phi(s) &= \pm\left(s - \log 2\sqrt{s} - \frac{1}{2} - \frac{3}{8s} + \cdots\right). \end{aligned} \tag{19}$$

And

$$\frac{\theta(s)}{\varphi(s)} = O(s^{-2})$$

From the form of the mapping function Φ and from the order of the integrand $\frac{\theta}{\Phi'}$, the relation (9) of hypothesis (V) is seen to be satisfied in this neighborhood, again with $\alpha = 0$, and uniformly in m, which is bounded.





At $s = 0$ the expansions are

$$\varphi(s) = \pm i s^{-\frac{1}{2}}\left(1 - \frac{s}{2} - \frac{s^2}{8} + \cdots\right),$$

$$\Phi(s) = \mp i\left(\frac{\pi}{2} - 2s^{\frac{1}{2}} + \frac{1}{3}s^{\frac{3}{2}} + \frac{1}{20}s^{\frac{5}{2}} + \cdots\right) \tag{20}$$

$$\theta(s) = \left(m^2 - \frac{1}{16}\right)s^{-2} + O(s^{-1})$$

And along the ordinary path which is chosen,

$$\int \left| \frac{\theta(s)}{\varphi(s)}\, ds \right| = \begin{cases} O(s^{-\frac{1}{2}}), & m \neq \pm\frac{1}{4} \\ O(s^{\frac{1}{2}}) & \text{in the particular case } m = \pm\frac{1}{4}. \end{cases}$$

In the first, general case,

$$s^{-\frac{1}{2}} = O(k^{\alpha}) \quad \text{if} \quad |s| > N\,|k|^{-2\alpha}.$$

If there be excluded from R_s a neighborhood of $s = 0$, of such shape, depending on $\arg \rho$ (or $\arg k$), as will not violate (IV), and whose dimensions are $O(k^{-2\alpha})$, the hypotheses are satisfied, again uniformly in m. The exponent r may be taken as $1 - \alpha$, which is positive.

In case m has the constant value $\pm\frac{1}{4}$, (9) is satisfied with $\alpha = 0$; no exception need be taken at $s = 0$. This leads to Weber's equation and the Hermite polynomials, for which these same formulas have been given by Schwid (loc. cit.).

4. **Asymptotic formulas for $W_{k,m}(z)$.** The conclusions of the theorem may now be asserted. There exist solutions of (1) having the forms (10) and (11) in R_s. To obtain asymptotic formulas for the particular function (3), it is necessary only to determine C_1 and C_2 as functions of k in a relation

$$W_{k,m}(z) = C_1 u_{p,1}(s, p) + C_2 u_{p,2}(s, \rho) \tag{21}$$

for some value of p. This may be done by use of the relation, seen from (5),

$$\lim_{s\to\infty} e^{\frac{s}{2}} z^{-k} W_{k,m}(z) = 1 \tag{22}$$

when

$$|\arg z| < \pi - \epsilon < \pi.$$

This limit will be attained in (21) by letting s become infinite.

If there be specified $\arg \Phi = \arg \varphi = 0$ when $\arg (s - 1) = 0$, then in (19) the upper and lower signs are to be taken respectively in the cases

$$\pi \geq \arg (s - 1) \geq -\pi,$$

and

$$3\pi \geq \arg (s - 1) \geq \pi.$$

And, as may be seen most readily from the figure and (19),

$$\arg \Phi \sim \begin{cases} \arg s \\ \arg s + 3\pi \text{ respectively,} \end{cases}$$

and

$$\arg \varphi \sim \begin{cases} 0 \\ \pi \text{ respectively.} \end{cases}$$

Now, by (19),

$$i\xi = i\rho\Phi = -2k\Phi = \pm\left(\frac{z}{2} - k\log z - k + k\log k + O\left(\frac{k}{s}\right)\right),$$

and

$$\begin{aligned} e^{\frac{s}{2}} z^{-k} &= e^{\frac{s}{2} - k\log s} \\ &= k^{-k} e^{k} e^{\mp i\xi}\left[1 + O\left(\frac{k}{s}\right)\right] \end{aligned} \tag{23}$$

In the first case, taking the upper signs,

$$\begin{aligned} \arg z &= \arg k + \arg s \\ &\sim \arg k + \arg \Phi \\ &\sim \arg \xi - \frac{\pi}{2}. \end{aligned}$$

Substituting (23), (10), and (21) with $p = 0$ or 1, into (22),

$$\lim_{s\to\infty} k^{-k} e^{k} e^{-i\xi} (2ik)^{-\frac{1}{4}} [C_1 e^{i\xi}\{1 + O(k^{-1})\} + C_2 e^{-i\xi}\{1 + O(k^{-1})\}] = 1,$$

or

$$\lim_{s\to\infty} k^{-k} e^{k} (2ik)^{-\frac{1}{4}} [C_1\{1 + O(k^{-1})\} + C_2 e^{-2i\xi}\{1 + O(k^{-1})\}] = 1,$$

when

$$\frac{3\pi}{2} - \epsilon > \arg \xi > -\frac{\pi}{2} + \epsilon. \tag{24}$$

If s becomes infinite in such a manner that $\epsilon < \arg \xi < \pi - \epsilon$, ξ in $\Xi^{(0)}$ or $\Xi^{(1)}$, $e^{-2i\xi}$ becomes infinite. Hence C_2 must be zero, for otherwise (24) is absurd. Then (24) yields

$$C_2 = 0, \qquad C_1 = k^{k} e^{-k} (2ik)^{\frac{1}{4}} [1 + O(k^{-1})]$$

where the restriction on arg k imposed by

$$-\pi < \arg s = \arg z - \arg k \leq \pi \tag{25}$$

serves to specify

$$k^k = |k|^k e^{ik \arg k}$$

and

$$(2ik)^{\frac{1}{2}} = |2k|^{\frac{1}{2}} e^{\frac{i}{2}\left(\frac{\pi}{2} + \arg k\right)}.$$

With these coefficients in (21),

$$W_{k,m}(z) = k^k e^{-k} \varphi^{-\frac{1}{2}} e^{i\xi}[1 + O(k^{-1}) + O(\xi^{-1})],$$

when $\quad -\pi + \epsilon \leq \arg \xi \leq 2\pi - \epsilon$, ξ in $\Xi^{(0)}$ or $\Xi^{(1)}$. (26a)

Similarly, using the lower sign in (23),

$$W_{k,m}(z) = k^k e^{-k} i\varphi^{-\frac{1}{2}} e^{-i\xi}[1 + O(k^{-1}) + O(\xi^{-1})],$$

when $\quad 2\pi + \epsilon \leq \arg \xi \leq 5\pi - \epsilon$, ξ in $\Xi^{(3)}$ or $\Xi^{(4)}$. (26b)

Using the formulas (11) and (12),

$$W_{k,m}(z) \sim k^k e^{-k} \varphi^{-\frac{1}{2}} (e^{i\xi} + ie^{-i\xi}),$$

when $\quad \pi + \epsilon \leq \arg \xi \leq 3\pi - \epsilon$, ξ in $\Xi^{(2)}$, (26c)

$$W_{k,m}(z) \sim k^k e^{-k} \varphi^{-\frac{1}{2}} (e^{i\xi} - ie^{-i\xi}),$$

when $\quad -2\pi + \epsilon \leq \arg \xi \leq -\epsilon$, ξ in $\Xi^{(-1)}$. (26d)

where only the principal terms have been written in (26c, d).

Further, from (13), if ξ is bounded, as it is if

$$s - 1 = O(k^{-\frac{1}{2}}),$$

or

$$z - 4k = O(k^{\frac{1}{2}}),$$

then

$$W_{k,m}(z) = k^k e^{-k} \left(\frac{4\pi k \Phi}{3\varphi}\right)^{\frac{1}{2}} \left\{e^{\frac{\pi i}{6}} J_{-\frac{1}{3}}(\xi) - e^{-\frac{\pi i}{6}} J_{\frac{1}{3}}(\xi)\right\} + k^k e^{-k} O(k^{-\frac{1}{2}}). \tag{27}$$

In terms of the original variables, z and k,

$$\varphi = \left(\frac{z - 4k}{z}\right)^{\frac{1}{2}}$$

and

$$i\xi = -\tfrac{1}{2}\sqrt{z(z-4k)} + 2k\log\left(\sqrt{z} + \sqrt{z-4k}\right) - k\log 4k.$$

The proper choices from among the several values of these functions may be reached most readily by continuation from the values when z, k, and $z - 4k$ are real and positive. Then, with $\arg \xi = \pi/2$ in $\Xi^{(1)}$, the radicals are all positive. The continuation is to be made within the bound imposed by

$$|\arg s| = |\arg z - \arg k| \leq \pi \tag{25}$$

The restriction $|s| > N|k|^{-2\alpha} = N|k|^{-2+2r}$ becomes

$$|z| > N|k|^{-1+2r},$$

in terms of r, the exponent in (26), $0 < r \leq 1$.

5. **Further results of Langer.** To obtain formulas asymptotically describing the Whittaker function in case z becomes zero with a greater rapidity than is permitted in the use of (26), it is necessary to reconsider equation (1) with the purpose of making an approximation which shall be uniformly valid in the neighborhood of $z = 0$. Langer [3] has given an extension of the theorem of section 1 which may be applied here.

The differential equation (1) has the form (6) with

$$\rho = k^{\frac{1}{2}}, \qquad \varphi(z) = z^{-1}, \qquad \chi(z) = \frac{\frac{1}{4} - m^2}{z^2} - \tfrac{1}{4}.$$

An approximation is

$$\frac{d^2y}{dz^2} + \left\{\frac{k}{z} + \frac{\frac{1}{4} - m^2}{z^2}\right\}y = 0 \tag{28}$$

which has solutions

$$y = \sqrt{kz}\,C_{2m}(2\sqrt{kz}) \tag{29}$$

for any Bessel function C of the order indicated.

Langer's results, of which the present example is a particularly simple application, consist of the fact that this approximation extends to the solutions, yielding here solutions of (1) of the forms:[12]

[12] The results of Langer [3] are extended as described in section 1. The integral $\int \left|\frac{\theta}{\varphi}dz\right|$ on p. 409 of Langer [3] reduces in the present example to $\frac{1}{2}\int |z^{\frac{1}{2}}dz|$ taken along a line through the origin and is $O(z^{3/2})$ rather than bounded. This has the effect of appending an extra factor $z^{3/2}$ in the O terms of (30) and (31) and replacing the restriction z bounded in (31) by $z = O(k^{1-\epsilon})$.

$$R(m) \geq 0; \qquad \arg k \text{ and } |kz| \text{ bounded},$$

$$u_1 = \sqrt{kz}J_{2m}(2\sqrt{kz}) + (kz)^{\frac{1}{2}+m}O(z^2)$$

$$u_2 = \sqrt{kz}Y_{2m}(2\sqrt{kz}) + \begin{cases} (kz)^{\frac{1}{2}-m}O(k^{-1}) & \text{if } m \neq 0 \\ (kz)^{\frac{1}{2}}\log(kz)O(k^{-1}) & \text{if } m = 0. \end{cases} \tag{30}$$

And if kz becomes infinite, $z = O(k^{1-\epsilon})$, these same solutions have the forms

$$\begin{aligned} u_1 &= \sqrt{kz}J_{2m}(2\sqrt{kz}) + k^{-1}z^{\frac{1}{2}}O(\exp|I(2\sqrt{kz})|), \\ u_2 &= \sqrt{kz}Y_{2m}(2\sqrt{kz}) + k^{-1}z^{\frac{1}{2}}O(\exp|I(2\sqrt{kz})|). \end{aligned} \tag{31}$$

6. **Formulas for $M_{k,m}(z)$.** The identification of the solution $M_{k,m}(z)$ in terms of the solutions u_1 and u_2 of the preceding section, as it is carried out here, involves a re-derivation of (30) by a comparison of the power series representing the functions $M_{k,m}(z)$ and $\sqrt{kz}J_{2m}(2\sqrt{kz})$.

With the notation

$$(\alpha)_r = (\alpha+1)(\alpha+2)\cdots(\alpha+r), \qquad (\alpha)_0 = 1,$$

(2) may be written

$$M_{k,m}(z) = z^{\frac{1}{2}+m}e^{-\frac{z}{2}}\sum_{r=0}^{\infty}\frac{(m-\frac{1}{2}-k)_r}{r!(2m)_r}z^r.$$

Now $e^{-\frac{z}{2}} = 1 - z/2 + O(z^2)$ and when $z = O(k^{-1})$ the sum in the preceding formula is bounded,

$$M_{k,m}(z) = z^{\frac{1}{2}+m}\left\{1 + \sum_{r=1}^{\infty}z^r\left(\frac{(m-\frac{1}{2}-k)_r}{r!(2m)_r} - \frac{(m-\frac{1}{2}-k)_{r-1}}{2(r-1)!(2m)_{r-1}}\right)\right\}$$
$$+ \Gamma(2m+1)O(z^{\frac{1}{2}+m})$$

$$= z^{\frac{1}{2}+m}\left\{1 + \sum_{r=1}^{\infty}z^r\frac{(m-\frac{1}{2}-k)_{r-1}\left(m-\frac{1}{2}-k+r-mr-\frac{r^2}{2}\right)}{r!(2m)_r}\right\}$$
$$+ \Gamma(2m+1)O(z^{\frac{1}{2}+m})$$

The numerator

$$P = (m-\tfrac{1}{2}-k)_{r-1}\left(m-\tfrac{1}{2}-k+r-mr-\frac{r^2}{2}\right)$$

is a polynomial in k with leading term $(-k)^r$ and second term zero. The sum of the absolute values of the coefficients is less than $Nr^3(|m|$

$+ \frac{1}{2})_r$ for some sufficiently large number N. Then $P = (-k)^r + P_1$ where $|P_1| < Nr^2(|m| + \frac{1}{2})_r |k|^{r-2}$.

From the expansion

$$\sqrt{kz}J_{2m}(2\sqrt{kz}) = \frac{(kz)^{\frac{1}{2}+m}}{\Gamma(2m+1)} \sum_{r=0}^{\infty} \frac{(-kz)^r}{r!(2m)_r}$$

it follows that

$$\frac{1}{\Gamma(2m+1)} M_{k,m}(z) = k^{-m} z^{\frac{1}{2}} J_{2m}(2\sqrt{kz}) + O(z^{\frac{1}{2}+m}) \tag{32}$$

provided only that $|kz|$ and $|m|$ are bounded.

Comparing (32) and the first equations of (30) and (31), which all hold in the range $N_1 < |kz| < N_2$,

$$\frac{1}{\Gamma(2m+1)} M_{k,m}(z) = k^{-m} z^{\frac{1}{2}} J_{2m}(2\sqrt{kz}) + k^{-(m+1)} z^{\frac{1}{2}} O(\exp |I(2\sqrt{kz})|) \tag{33}$$

when $\arg z$ and $\arg k$ are bounded, whether $R(m)$ be positive or negative, $z = O(k^{1-\epsilon})$.

To carry out the description in terms of elementary functions, the asymptotic evaluation of the function J_{2m}, given in the following convenient form by Langer, may be used.

$$J_{2m}(2\sqrt{kz}) = \sqrt{2}(kz)^{\frac{1}{4}} \{c_1 e^{2i\sqrt{kz}} + c_2 e^{-2i\sqrt{kz}} + (kz)^{-\frac{1}{4}} O(\exp |I(2\sqrt{kz})|)\} \tag{34}$$

where

$$c_1 = \frac{1}{\sqrt{2\pi}} e^{(4s-1)(4m+1)\frac{\pi i}{4}},$$

when $(2s-2)\pi + \epsilon < \arg \sqrt{kz} < (2s+1)\pi - \epsilon$,

$$c_2 = \frac{1}{\sqrt{2\pi}} e^{(4s+1)(4m+1)\frac{\pi i}{4}},$$

when $(2s-1)\pi + \epsilon < \arg \sqrt{kz} < (2s+2)\pi - \epsilon$.

There is, for any value of $\arg \sqrt{kz}$ not a multiple of π, a solution of (1) which, when evaluated by (33) and (34), contains only that one of the two terms $e^{\pm 2i\sqrt{kz}}$ for which the real part of the exponent is negative. In this situation the exponent in the O term is also to be made negative: $O(\exp - |I(2\sqrt{kz})|)$ (see Langer [3], the subdominant solution).

The formula with (34) may be shown to be in complete agreement with (26) in the range $N_1 \mid k \mid^{-1+\epsilon_1} < \mid z \mid < N_2 \mid k \mid^{\frac{1}{3}-\epsilon_2}$ in which both are valid.

7. **The Laguerre Polynomials.** The asymptotic formulas as $n \to \infty$ for the Laguerre polynomials, given by (5), may now be reached without difficulty by specializing the formulas of the preceding sections. The following evaluations give a complete description for $z = x$ real and positive.

From (32), when $x < cn^{-1}$, c a positive constant,

$$L_n^{(\alpha)}(x) = \frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)} x^{-\frac{\alpha+1}{2}} e^{\frac{x}{2}} \left[\left(n+\frac{\alpha+1}{2}\right)^{-\frac{\alpha}{2}} x^{\frac{1}{2}} J_\alpha\left(2\sqrt{\left(n+\frac{\alpha+1}{2}\right)x}\right) + O\left(x^{\frac{5+\alpha}{2}}\right)\right]. \tag{35}$$

From (33), when $c_1 n^{-1} < x < c_2 n^{\frac{1}{3}-\epsilon}$,

$$L_n^{(\alpha)}(x) = \frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)} x^{-\frac{\alpha+1}{2}} e^{\frac{x}{2}} \left[\left(n+\frac{\alpha+1}{2}\right)^{-\frac{\alpha}{2}} x^{\frac{1}{4}} J_\alpha\left(2\sqrt{\left(n+\frac{\alpha+1}{2}\right)x}\right) + O\left(x^{\frac{1}{4}} n^{-\frac{\alpha}{2}-\frac{3}{4}}\right)\right] \tag{36}$$

$$= n^{\frac{\alpha}{2}}\left(1+\frac{\alpha^2-3\alpha}{n}\right) x^{-\frac{\alpha}{2}} e^{\frac{x}{2}} J_\alpha\left(2\sqrt{\left(n+\frac{\alpha+1}{2}\right)x}\right) + O\left(e^{\frac{x}{2}} x^{-\frac{\alpha}{2}+\frac{5}{4}} n^{\frac{\alpha}{2}-\frac{3}{4}}\right) \tag{37}$$

$$= \pi^{-1} n^{\frac{\alpha}{2}-\frac{1}{4}} e^{\frac{x}{2}} x^{-\frac{\alpha}{2}-\frac{1}{4}} \left[\cos\left(2\sqrt{nx} - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + O(n^{-\frac{1}{4}} x^{\frac{1}{4}}) + O(n^{-\frac{1}{4}} x^{-\frac{1}{4}})\right]. \tag{38}$$

From (26c) with $\arg \xi = 2\pi$, when $c_1 n^{-1+2r} < x < 4n - c_2 n^{\frac{1}{3}}$, $1 \geq r > 0$,

$$L_n^{(\alpha)}(x) = (-1)^n \left(\frac{2}{\pi}\right)^{\frac{1}{2}} n^{\frac{\alpha}{2}} x^{-\frac{\alpha}{2}-\frac{1}{4}} e^{\frac{x}{2}} (4n+2\alpha+2-x)^{-\frac{1}{4}} \left[\cos\left\{\tfrac{1}{2}\sqrt{x(4n+2\alpha+2-x)} - (2n+\alpha+1)\right.\right. \tag{39}$$

$$\left.\left.\cos^{-1}\sqrt{\frac{x}{4n+2\alpha+2}} + \frac{\pi}{4}\right\} + O(n^{-1}) + O(n^{\frac{1}{3}}(4n-x)^{-\frac{1}{2}})\right].$$

From (26a) with $\arg \xi = \pi$, when $4n + cn^{\frac{1}{3}} < x$,

$$L_n^{(\alpha)}(x) = x^{\frac{1}{4}}(x - 4n - 2\alpha - 2)^{-\frac{1}{4}} e^{-\frac{1}{2}\sqrt{x(x-4n-2\alpha-2)}} \left(\frac{\sqrt{x} + \sqrt{x - 4n - 2\alpha - 2}}{2\sqrt{e}}\right)^{2n+\alpha+1} \tag{40}$$

$$[1 + O(n^{-\frac{1}{2}}) + O(n^{\frac{1}{2}}(x - 4n)^{-\frac{1}{2}})].$$

And substituting in (27), there may be obtained the formulas for the remaining case, $4n - c_1 n^{\frac{1}{3}} < x < 4n + c_2 n^{\frac{1}{3}}$.

UNIVERSITY OF CINCINNATI,
CINCINNATI, OHIO.

THE LIGHT FIELD

By A. Gershun

Translated by Parry Moon and Gregory Timoshenko

Theoretical photometry constitutes a case of "arrested development", and has remained basically unchanged since 1760 while the rest of physics has swept triumphantly ahead. In recent years, however, the increasing needs of modern lighting technique have made the absurdly antiquated concepts of traditional photometric theory more and more untenable.

A vigorous attempt to bring the theory of light calculation into conformity with the spirit of physics is now in progress. Professor Gershun of the State Optical Institute, Leningrad, is one of the pioneers in this reform, some of his work being contained in the book, *"The Light Field"**, a translation of which follows.

The purpose of our translation is two-fold: first, to bring to the attention of engineers the new methods of radiometric and photometric calculation; and secondly, to interest mathematicians and physicists in the further development of this branch of field theory. In its present form, the concept of the light vector is of distinct practical value and its general use by illuminating engineers would tend to simplify calculations and clarify thinking.

Valuable as the methods are, however, they probably do not constitute the ultimate solution of the problem. The light field considered in this book is a classical three-dimensional vector field. But the physically important quantity is actually the *illumination*, which is a function of *five* independent variables, not three. Is it not possible that a more satisfactory theory of the light field could be evolved by use of modern tensor methods in a five-dimensional manifold? We must look to the mathematician for any such development.

* *Svetovoe pole*, Moscow, 1936. A brief treatment is also given by Gershun in a paper, Notions du champ lumineux et son application á la photométrie. R. G. E., **42**, 1937, p. 5. The paper is criticized by A. Blondel, Sur une prétendue théorie du champ lumineux. R. G. E., **42**, 1937, p. 579. We have weighed Blondel's criticisms and do not consider that they vitiate Gershun's results. The reader will find the elements of light-field theory also in Chap. X of Moon, Scientific Basis of Illuminating Engineering, New York, 1936.

COLE MEMORIAL LIBRARY

The translation aims at an accurate reproduction of the spirit of the original, but it is not a literal translation. Sentences and even paragraphs have been omitted at times to fit the work for a somewhat different class of readers than that for which the book was originally written. No additional material has been added, with the exception of a few footnotes, which are inclosed in square brackets. We have taken the liberty, however, of numbering the equations and the references.

THE TRANSLATORS,
Massachusetts Institute of Technology.

CONTENTS

CHAPTER
I. Introduction 53
II. Fundamental Theory of the Light Field 58
III. Space Illumination 77
IV. The Light Vector 90
V. The Structure of the Light Field 107
VI. Photometric Computations 127
VII. Mathematical Theory 141

CHAPTER I

INTRODUCTION

1. Problems of Theoretical Photometry

Let us define first the meaning of the word *photometry*. Customarily, the term is given a rather narrow meaning corresponding to a literal translation from the Greek (φῶς—light, μέτρον—measure). Photometry, in this sense, is that part of optics dealing with the measurement of radiant energy, evaluated according to its visual effect and related therefore solely to the visible part of the spectrum. Such a definition has historical sanction, since for many years the eye was the sole instrument used in evaluating radiant energy. This purely physiological criterion, as applied directly to basic measurements of quantities characterizing the distribution of radiant energy, resulted in a peculiar set of photometric concepts, standards, and units which stood apart from the general system of physical units. The usual treatise on photometry consisted largely in descriptions of apparatus used in the visual comparison of radiations.

In the present treatment, however, we attach a rather broad meaning to the words *light* and *measurement* when we define photometry as a branch of science dealing with the measurement of light. In "light" we include all radiations, visible or invisible. This corresponds, for example, to the definition of photometry given by Charles Fabry, who says[1] that photometry is "la mesure de l'intensite des rayonnements, visibles ou non visibles, quelle que soit leur place dans le spectre et quel que soit l'appareil de mesure employé."

The word *measurement* is also defined in a broad sense, since it is not confined to the experimental technique but is used rather for the entirety of theoretical and experimental questions connected with quantitative comparisons. Thus photometry deals with energy relationships in emission processes and in the propagation and absorption of radiation. The radiation may be either visible or invisible. The quantity of radiation may be evaluated in units of energy or in its effect upon a receptor: the human eye, the photographic plate, the human skin, etc. Depending on the receptor, the result will be evaluated either in the usual physical units or in some special units, as light

[1] Charles Fabry: Introduction générale a la photométrie, Paris, 1927.

units, photographic units, or erythemal units. In these cases, to each monochromatic component of radiation is attributed a weight-number, corresponding to the response produced in the receptor by a unit quantity of radiant energy of the given wavelength.

All theorems of photometry are independent of the units in which the quantity of radiation is measured (energy units, for example, or photometric units). For the sake of convenience, *photometric terminology is used in this book.**

Before treating the more recent problems of theoretical photometry, we shall give a brief account of the history of its development. Photometry was founded as a scientific discipline by a French scientist, Pierre Bouguer (1698–1758),[2] who was the author also of researches in mathematics, astronomy, geodesy, physics, geophysics, and navigation. Johann Heinrich Lambert (1728–1777) gave a mathematical form to some of the ideas originated by Bouguer and elaborated upon methods of photometric computation.[3] In the modest history of photometry, Bouguer and Lambert deserve a position corresponding to that occupied in the history of electromagnetic phenomena by those Titans of science, Faraday and Maxwell. The importance of the work of Bouguer and Lambert has become evident only in recent times, when the questions considered by them have become of vital interest to engineers.†

In the one hundred and fifty years following the work of the founders of photometry, the questions of photometric computations were neglected almost completely.[4] The problems of theoretical photometry were pushed aside from the main path of the development of physics. This resulted in the present state of photometric theory, based upon eighteenth-century mathematics and devoid of generality. The con-

* [Though all the theorems can be interpreted equally well in radiometric terms.] The Translators.

[2] Traité d'optique sur la gradation de la lumière. Ouvrage posthume de M. Bouguer, de l'Académie Royale des Sciences, Paris, 1760. A less complete edition of work by Bouguer appeared under the title "L'Essai d'Optique" in 1729.

The first book on photometry was written by a Paris Capuchin, François Marie: Nouvelles découvertes sur la lumière, Paris, 1700.

[3] J. H. Lambert, Photometria sive de mensura et gradibus luminis, colorum et umbrae, 1760. German edition with annotations by E. Anding: Ostwald's Klassiker der exacten Wissenschaften, Nos. 31–33, Leipzig, 1892.

† Note that the problems of photometric computations are of interest not only in optics and illuminating engineering but also in other branches of science, as for example in heat engineering.

[4] See, however, August Beer, Grundriss des photometrischen Calcüles. Braunschweig, 1854.

cepts of the force field and the methods of vector analysis are absent. This condition can be explained by the fact that the structure of the theory has not been of special interest to physicists and has not been needed until recently by engineers.

Problems of theoretical photometry have come into prominence with the growth of illuminating engineering. This growth has, in turn, been caused by the demands of humanity for better artificial lighting—demands that could be satisfied for the first time by means of electrical light sources. For some time the engineer required from the theory only methods of calculating illumination from point sources, but the recent development of luminous architectural elements and of methods of calculating daylighting in buildings have necessitated the study of large surface sources. Problems of reflector design have been developed also. Knowledge of the reflecting properties of specular surfaces and diffusing surfaces was required, as well as knowledge of the transmission of light through absorbing and diffusing media. Thus appeared a series of practical problems requiring a large variety of photometric computations and giving the first impulse toward the construction of a *generalized theory of photometry*.

In the first quarter of this century, illuminating engineers were concerned primarily with the problem of obtaining the required illumination on the working surfaces. Experience showed this criterion to be quite inadequate, however, since the illumination of a working surface is not a universal measure of the lighting. In everyday design practice, consideration must be given to the magnitude of the illumination and also to the correct coordination of general and local illumination, to the direction of the light, and to shadows. In illuminating engineering literature one finds the expression "quality of lighting", by which is understood all the properties that the engineer cannot characterize as yet by definite numbers. The problem of the quality of lighting has been responsible for the introduction of several new concepts in illuminating engineering: the *space illumination* which characterizes the density of light at a point in space, independent of the light direction; and the *light vector*, which characterizes the predominant direction of the light. Illuminating engineering, which was originally merely a branch of electrical engineering, has become an independent and broad discipline dealing with theoretical methods and technical applications of radiant energy.

Illuminating engineering has need of a theoretical basis, analogous to the theory of the electromagnetic field which forms the foundation of

electrical engineering. In the author's opinion, the *theory of the light field* must form one of the most important parts of this theoretical foundation.

2. The Light Field

Problems of theoretical photometry fit nicely into the frame of classical mathematical physics. Mathematical physics includes the analytical treatment of heat conduction, elasticity, electricity and magnetism. All these branches are united by a general theory of the physical field, which can be classified as physics or as geometry. We may define the physical field as a part of space, studied from the standpoint of a definite physical process happening within that space. Analogously, we shall introduce the concept of the *light field*, as a part of space studied from the standpoint of transmission of radiant energy within that space. Until recent times, photometry limited itself to concepts concerning the emission and absorption of light by bodies, while the transmitting medium was ignored. The older photometric science was a peculiar manifestation of the concept of *actio in distans*. The modern study of the light field consists of an investigation of the space-distribution of luminous flux. The separate photons are disregarded and the assumption is made that radiant energy is continuous in time and in space and that the flux of this radiant energy varies continuously from point to point. The familiar methods of vector analysis are used in the theoretical study of the light field.

A physicist may ask why the author distinguishes the light field from the well-known electromagnetic field. Obviously, the light field is caused by electromagnetic phenomena, but it is quite different in its qualities from the electromagnetic field. In studying the latter we are considering the electric and magnetic forces caused by an elementary emitter. In studying the light field, however, we deal with bodies of finite size consisting of great numbers of elementary emitters. In contrast with the elementary electromagnetic field, we have here a macrocosmos with respect to time as well as space. We assume that the differential of time exceeds by far the period of a single vibration and that the differential of distance is far greater than a single wavelength.[5] The theory, as presented in this book, corresponds to *geometrical optics* rather than physical optics. We shall deal with a beam of light which propagates in a straight line through a homogeneous

[5] Max Planck, Vorlesungen über die Theorie der Wärmestrahlung. Leipzig, 1923.

medium, and which is the carrier of radiant energy. Thus the treatment will be a geometrical one, to which is added the concept of energy. The development of a wave-theory of the light field may be considered as a step to follow. One feels the need of this extension, for example, in the absolute determination of the energy distribution in the optical image.

CHAPTER II

Fundamental Theory of the Light Field

1. Fundamental Photometric Quantities

We shall consider first the quantities used in the study of the light field. It is most convenient to choose as fundamental magnitude the amount of radiant energy transferred through a hypothetical surface per unit time. This quantity is called *radiant flux*. However, we shall use the term **luminous flux**, the photometric analog of radiant flux, in accordance with our convention of employing the names associated with photometric concepts rather than with radiometric concepts. The book is written for physicists, however, as well as for illuminating engineers, and the derivations presented here will apply to radiant flux as well as to luminous flux.

Let us first recall the definitions of the photometric quantities: *luminosity*, *illumination*, *intensity*, and *brightness*.* Consider a point O of the surface S of a luminous body in space (Fig. 1). The element of surface dS containing the point O is chosen sufficiently small so that it may be considered uniformly bright over its surface. The total luminous flux emitted in all directions from the element dS is denoted by dF and the **luminosity** L of the surface S at the point O is defined as

$$L = \frac{dF}{dS} \tag{1}$$

Luminosity is surface density of emitted luminous flux. The total flux emitted by the source is

$$F = \int_S L dS \tag{1a}$$

Now consider an element ds, illuminated by the luminous surface S. The **illumination** E of the surface s at the point P is defined as

$$E = \frac{dF'}{ds} \tag{2}$$

where dF' is the total luminous flux on ds from all directions of space. *Illumination is surface density of received luminous flux.*

* [Svetimost, osveshchennost, sila sveta, and yarkost]. Translators.

It is obvious that the concept of illumination may be divorced from the concept of an actual illuminated body. One may define illumination at a point in the light field, on an arbitrarily oriented surface, by using the abstract concept of a geometric element of surface placed at the point in question. This concept, which is characteristic of physical field theory, may already be found in illuminating engineering, e.g. in the expression, "the horizontal illumination at the elevation of one meter from the floor". Since the orientation of the surface element is completely and uniquely defined by the direction of the surface normal, the illumination may be regarded as a function of two factors: position in space, and direction.

In computing the values of illumination, we must know the intensity* of each element of the surface of the luminous body in all directions. Consider the regions of the light field that are remote from the luminous body S, so that the dimensions of the source are negligible in comparison with the distance from it to the point P. Then the body S can be considered as a point source. For a study of these distant parts of the light field, it is sufficient to use the concept of the **luminous intensity** of the entire body in the given direction. This concept is derived in the following manner. Consider a definite elementary bundle of rays diverging from the source. As a measure of the multitude of these rays, we shall use the solid angle $d\Omega$ corresponding to the conical surface enclosing all the rays. The luminous flux in these directions has the value dF. The luminous intensity I in the given direction is defined as

$$I = \frac{dF}{d\Omega} \tag{3}$$

When the values of the intensity in various directions are known, the light field of a luminous body is defined at a sufficiently large distance from it. But the concept of intensity loses its meaning if the solid angle becomes large; and it is then necessary to divide the radiating surface into small elements, each of which is considered as a point source.

Let us recall the fundamental laws defining the light field produced by an element of luminous surface. These laws follow from purely geometrical considerations, together with the law of conservation of energy. If dI represents the intensity of the element dS in the direction of the illuminated element ds (Fig. 1), and if the distance between

* [Or candlepower]. *Trans.*

the two elements is far greater than the dimensions of either of them, the illuminated element ds receives from dS the flux d^2F, which is an infinitesimal part of the flux dF emitted by the element dS. The index 2 denotes an infinitesimal of higher order. The luminous flux is

$$d^2F = dI\, d\Omega \tag{4}$$

where $d\Omega$ is the solid angle defined by ds. Let r be the distance between ds and dS, and let θ be the angle between the incident rays and the normal n of the illuminated element ds. Then

$$d\Omega = \frac{ds \cos\theta}{r^2} \tag{5}$$

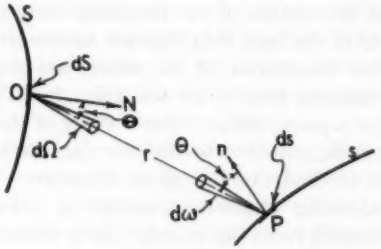


FIG. 1

and the illumination dE caused by radiation from the element dS is

$$dE = \frac{d^2F}{ds} = \frac{dI}{r^2}\cos\theta \tag{6}$$

This fundamental photometric formula expresses two relations: the inverse-square law and the cosine law of illumination. These elementary relations may be applied also in the case of a source of finite size at distances sufficiently large compared to the size of the body.

Let us now introduce the concept of **brightness** B of a luminous surface S at a point O, in the direction OP. Denote by Θ the angle between the direction OP and the normal N of the luminous element dS, and by $d\sigma$ the area of the projection of the element dS upon a plane normal to OP. Then

$$B = \frac{dI}{d\sigma} \tag{7}$$

where

$$d\sigma = dS \cos\Theta$$

The flux d^2F may be expressed as

$$d^2F = \frac{B \cos\theta \cos\Theta \, dS \, ds}{r^2}$$

But

$$d\omega = \frac{dS \cos\theta}{r^2} \tag{8}$$

Thus,

$$dE = B \cos\theta \, d\omega. \tag{9}$$

This relation represents another form of the basic law of photometric theory. Denoting by dE_n the normal component of the illumination at the point P, we obtain

$$dE_n = B \, d\omega \tag{10}$$

and

$$B = \frac{dE_n}{d\omega} \tag{10a}$$

Analogous to the definition of intensity, which is measured by the luminous flux per unit solid angle, the *brightness of a luminous source may be defined as the normal component of illumination produced by that source per unit solid angle.* This relation may be considered as a fundamental definition of brightness, which must be given preference over the usual definition of brightness as intensity per unit area. The latter can be used only in the special case of a light-emitting surface, and fails, for example, in evaluating the brightness of the sky. The concept of sky brightness is indeed a puzzling one when the classical idea of brightness (candlepower per square centimeter of the surface of the source) is used. It is difficult to decide where to locate these square centimeters which have to be divided into the equally illusive luminous intensity.

2. Brightness of a Ray

The concept of brightness must now be generalized to adapt it for use in the theory of the light field. Consider the light field produced by a number of luminous bodies S_1, S_2, $\cdots$ (Fig. 2), and let us characterize the structure of the light field at an arbitrary point P. Consider a plane s through the point P, and assume this plane to be opaque except for a small area *ds* containing the point P. This area is sufficiently small so that the light field may be considered uniform at all points of *ds*, but it is sufficiently large so that diffraction phenomena may be

neglected. On this assumption, the fluxes dF_1 and dF_2 are obviously proportional to the size of the opening ds, or

$$\left.\begin{aligned} dF_1 &= E_1\,ds \\ dF_2 &= E_2\,ds \end{aligned}\right\} \tag{11}$$

The coefficients of proportionality E_1 and E_2 represent the surface density of luminous flux on the given side of the aperture, or E_1 and E_2 represent the illuminations of the two sides of the plane s at the point P. E_1 is the illumination from side 1 and E_2 is the illumination from side 2. (Fig. 2.) For computation of the energy balance we shall assume

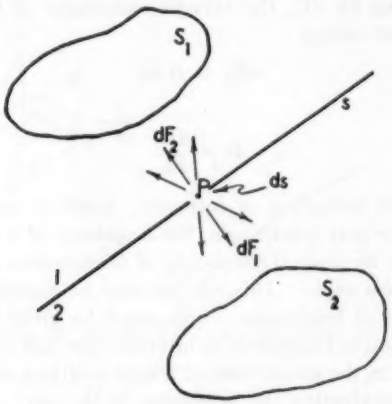


FIG. 2

$dF_1 > dF_2$, which means that the energy is transmitted through the diaphragm ds, entering through the side 1 and emerging through the side 2. The quantity of light transmitted per unit time is equal to $dF = dF_1 - dF_2$. The surface density of the luminous flux through the aperture is

$$D_s = \frac{dF}{ds} = E_1 - E_2. \tag{12}$$

The quantity D_s determines the transfer (flow) of energy through unit area of the aperture. We shall call D_s the **flux density**.

The aperture ds, through which the light passes, may be considered as an independent source of light, and we may apply to it all the con-

cepts used in the study of light sources. At great distances, the aperture may be considered as a point source with a definite intensity in each direction. We may also introduce the concept of brightness of the aperture in various directions, defining this quantity as was done in considering the brightness of real surfaces. Introducing this concept of brightness allows one to derive an energy balance at the aperture ds and also to characterize the directional distribution of the luminous flux through the aperture.

Consider a direction P_1P_2 making an angle θ_1 with the normal to the screen s_1 (Fig. 3), and consider a solid angle containing this direction P_1P_2. The solid angle may be constructed in the following manner: we make use of a second opaque screen s_2 with an infinitely small aperture ds_2 containing the point P_2. The elementary solid angle $d\omega_1$ is defined by a conical surface with its apex at P_1 and having a base ds_2.

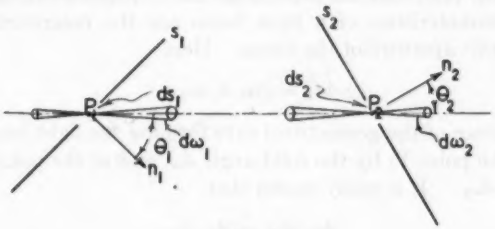


FIG. 3

This solid angle is chosen sufficiently small so that all directions contained within it may be considered equal. The luminous flux d^2F through the aperture ds_1 and enclosed by the angle $d\omega_1$ is

$$d^2F = dI\, d\omega_1 \tag{13}$$

The coefficient of proportionality dI represents the luminous intensity of the aperture ds_1 in the direction P_1P_2. This quantity depends on the size and on the orientation of the element ds_1 and is proportional to the area

$$d\sigma_1 = \cos\theta_1\, ds_1 . \tag{14}$$

Referring the intensity to a unit area of the projection; that is, determining the brightness of the aperture ds_1 in the direction P_1P_2, we obtain a fundamental photometric concept which depends solely on the position of the point P_1 and on the direction P_1P_2. This quantity is

called *the brightness B at the point P_1 in the direction P_1P_2*. To each point in the light field and to each direction at that point corresponds a definite value of brightness. In other words, brightness is a function of position and of direction. One may say that B represents the brightness of the light ray P_1P_2 at the point P_1. The concept of the brightness of a light ray was first introduced by Planck.

In geometrical optics, the light ray is identical with a straight line and has no thickness. But obviously energy cannot be propagated through zero cross-sectional area. In photometry we must use a light beam composed of an infinite number of adjacent geometrical rays, and an infinitesimal amount of energy is carried by such a beam. The light beam is determined in this case by two infinitesimal apertures ds_1 and ds_2, which are at a finite distance from each other. The geometry of the light beam is characterized at the input aperture by the area $d\sigma_1$, Eq. (14), and the solid angle $d\omega_1$. Equivalent to these geometrical characteristics of a light beam are the magnitudes $d\sigma_2$ and $d\omega_2$ at the exit aperture of the beam. Here

$$d\sigma_2 = \cos \theta_2 \, ds_2 \tag{14a}$$

The divergence of the geometrical rays forming the light beam is determined at the point P_1 by the solid angle $d\omega_1$ and at the point P_2 by the solid angle $d\omega_2$. It is easily shown that

$$d\omega_2/d\omega_1 = d\sigma_1/d\sigma_2 \tag{15}$$

so

$$d\sigma_1 \, d\omega_1 = d\sigma_2 \, d\omega_2 \tag{16}$$

A more rigorous proof of Eq. (16) is given in Section 1, Chap. VII.

Let us consider the variation of brightness along the light beam. The brightness at the point P_1 is denoted by B_1, while that at P_2 is denoted by B_2. The flux d^2F through the aperture ds_1, in the direction ds_1–ds_2, is

$$d^2F_1 = B_1 \, d\sigma_1 \, d\omega_1 \tag{17}$$

which represents the luminous flux in the light beam at the point P_1. Similarly, at P_2,

$$d^2F_2 = B_2 \, d\sigma_2 \, d\omega_2 \tag{18}$$

Taking into account that $d\sigma \, d\omega$ is constant, we find that the brightness along the ray is directly proportional to the luminous flux, or

$$B_2 : B_1 = d^2F_2 : d^2F_1 \tag{19}$$

If the light beam emerges into empty space, then, by the principle of conservation of energy,

$$d^2F_2 = d^2F_1$$

and

$$B_1 = B_2 \tag{20}$$

or the brightness is the same at all points of the light beam. This conclusion is often formulated in text books of physics as the principle of independence of the brightness of a luminous body with distance, and the principle is often used in the solution of problems of illuminating engineering where absorption and dispersion of light can be neglected. This explains why the concept of the brightness in a beam is not widely used in illuminating engineering; for if brightness is the same at all points in a beam, it is simpler to talk about the brightness of the source in the given direction, rather than the brightness in the beam. Thus

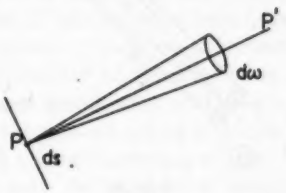


FIG. 4

the concept of brightness is customarily associated intimately with luminous sources. But if in the path of the light beam there are regions which emit, absorb, or disperse the light, then the luminous flux carried by the beam changes from point to point and the brightness varies correspondingly. For example, if the light beam passes through an absorbing medium, the luminous flux decreases because of the gradual change of radiant energy into heat energy. The luminous flux decreases exponentially and the brightness of the beam varies according to the same law.

The necessity of introducing the concept of brightness as a function of position and of direction may be illustrated by the question of the brightness of the sky, which depends on the direction in which it is observed as well as the position of the observer. Thus brightness should be considered as a function of position and direction in the light field. Let us place at a given point a small plane element *ds* which is perpendicular to the direction PP′ (Fig. 4). The brightness is considered

as the coefficient by which the solid angle $d\omega$ must be multiplied to obtain the illumination dE_n produced on the element ds by the luminous flux in the angle $d\omega$, or

$$B = dE_n/d\omega \tag{21}$$

As in the measurement of illumination, brightness may be measured objectively by means of a photocell. The photocell is mounted on an opaque tube m (Fig. 5) which intercepts all the rays not contained within the solid angle ω. If the length of the tube is considerably larger than the area of the photocell the luminous flux on the photocell may be considered as proportional to the average value of the brightness within the solid angle ω.

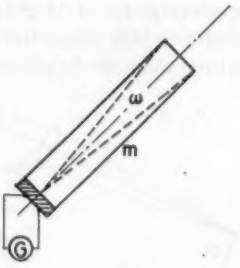


FIG. 5

3. The Brightness-distribution Solid

The structure of the light field at a point P may be studied in the following manner. Let us form an infinite number of infinitesimal solid angles adjoining each other and filling all space, all the solid angles having a common origin at P. The transfer of light within an arbitrary solid angle $d\omega$ is characterized by the illumination dE_n of a surface element placed at that point, with its normal collinear with the axis of the solid angle $d\omega$. The values dE_n depend not only on the character of the light field but also on the initial arbitrary subdivision of the space. Dividing the values of dE_n by the corresponding values $d\omega$, however, we obtain values of brightness which are functions of direction and which are independent of the original geometrical subdivision. The totality of these values of brightness determines completely the structure of the light field at the point P.

The distribution of brightness may be characterized geometrically in the following manner. Let us mark off in each direction from the

point P a distance which represents the brightness in that direction. The geometric locus of the ends of these distances forms a closed surface. This surface may be called the *brightness-distribution surface* for the given point in the light field. The space enclosed by this surface is called the *brightness-distribution solid*. The section of the *brightness-distribution surface* by a plane passing through P gives in polar coordinates the *brightness-distribution curve* in that plane. Thus for completely diffused light, where brightness is the same in all directions, the brightness-distribution solid is represented by a sphere with center at P. For a parallel beam of light where the brightness is zero in all directions except one, the brightness-distribution solid degenerates into a line segment in the direction of the beam.

Depending on the values of brightness in various directions, the eye evaluates the level of illumination of an entire room and also evaluates the uniformity of the lighting, the degree of diffusion, and the presence or absence or glare. It seems peculiar that until recent times the quality and quantity of lighting have been judged by the distribution of illumination at various points on a horizontal plane. The visual evaluation of the uniformity of illumination is determined in the first place by the brightness distribution of the surfaces enclosing the room. This was first correctly pointed out by L. Bloch.[6]

For example, with daylight entering a room from the side, the horizontal illumination near the windows is many times as great as the illumination at points distant from the windows. The illumination of such a room is highly nonuniform, according to present methods of evaluation. We know, however, that if the room is not too deep, and is illuminated by windows in one of the walls, it appears to be illuminated sufficiently uniformly. This can be explained by the fact that there is almost equal illumination on the side and back walls; for although the back wall is further from the windows, the light rays are normal to it. Even if the eye is fixed on a definite work surface, the distribution of brightness in all other directions plays an important rôle.

4. Functions of Position and Direction

The brightness-distribution solid characterizes completely and uniquely the distribution of luminous flux at any point of the light field. Knowing the brightness-distribution solid, one can compute the

[6] L. Bloch, Die Leuchtdichteverteilung im Raum. Licht und Lampe, 1930, No. 13, p. 663.

values of all other magnitudes serving to describe the light field at the given point. Thus, for example, the brightness-distribution solid allows one to find the illumination of an arbitrary oriented element ds at a given point P (Fig. 1). Let us denote by B the brightness in the direction of $d\omega$. The illumination of the element ds by the rays contained within the solid angle $d\omega$ is

$$dE = \cos\theta \, dE_n \tag{22}$$

where

$$dE_n = B \, d\omega$$

is the normal component of illumination. The total illumination of the element ds is

$$E = \int_{(2\pi)} \cos\theta \, dE_n = \int_{(2\pi)} B \cos\theta \, d\omega \tag{23}$$

where the sign (2π) indicates that the integration is performed over a solid angle 2π, or all directions on one side of the element ds.

For convenience, let us introduce some terms and notations that will allow us to describe more briefly the characteristics of the light field. Thus, the expression "illumination of the element ds placed at the point P and having its normal in the direction P_n" may be shortened to "illumination at the point P in the direction P_n". The illumination in a given direction can be represented by an arrow oriented in that direction, the arrow tail being placed at the point in question and its length being equal to the numerical value of the illumination. If an element of surface is illuminated from both sides, it is necessary to agree on the *sense* of the normal. Some authors use the outward-drawn normal to the illuminated surface, while others use the inward-drawn normal. This question is considered by Yamauti in his paper on the theory of the light field.[7] In illuminating engineering practice it is customary to employ the outer normal to the illuminated surface; but in theoretical treatments it seems more natural to choose the opposite direction, for then the illumination is represented by an arrow directed in the sense of the incident light. A similar question arises in constructing a graph of brightness distribution in various directions. We shall not bind ourselves to definite rules but shall apply in each separate case the method that is most convenient.

[7] Ziro Yamauti, Theory of Field of Illumination. Researches of the Electrotechnical Laboratory, Tokyo, No. 339, 1932.

Luminous intensity, brightness, and illumination can be represented graphically by arrows at the point in question. One might conclude that intensity and brightness are vector quantities. However, this is wrong. Not every magnitude requiring for its determination a number and a direction represents a vector, though the vector is often defined in such a manner. The vector concept contains the principle of geometric addition. Intensity and brightness are not vectors, inasmuch as the vectorial addition of intensities or of brightnesses in different directions has no meaning. Two vectors representing the same kind of quantity at a point in space can usually be replaced by a single vector found by the parallelogram rule. But the brightnesses of beams arriving at a point from various directions are completely independent of each other.

One must approach the vector concept of photometric quantities in the following manner. Each physical field is characterized by a number of quantities. The totality of values of each of these magnitudes for all points of the field (and in some cases also for all directions at each of these points) may be defined as the field of that quantity. Let us note that the concepts of the physical field and of the field of each of these quantities are just as different in principle as the physical process itself and those mathematical relations that describe it. The quantities characterizing the light field may be subdivided in two fundamental classes:

(1) functions of position and direction
(2) functions of position

The latter class may again be subdivided into two groups:

(a) vector point-functions
(b) scalar point-functions.

A function of position and direction may be defined in the following manner. Let us take a region of space and assume that to each point in this region and to each direction at this point corresponds a definite numerical value of some quantity. We shall call this magnitude a function of position and direction. As an example of the first class of quantities we have brightness and illumination. Luminous intensity may also be put in this class. Such a quantity (more exactly speaking, its value for a given direction) may be represented by an arrow in space. However, the function of position and direction is not a vector in any sense. A vector at a given point is defined by a single arrow; a function of position and direction is defined by a bundle of arrows, one for each direction of space.

A quantity of prime importance in the study of the light field is the *flux density D*, that is, the difference in illumination of the two sides of an infinitesimal diaphragm. This quantity characterizes the transfer of light through a unit area. It is a function of the position of the diaphragm and of the direction (normal to the surface of the diaphragm) and it may be regarded, as it will be shown later, as a projection of a vector. This vector is an example of a vector point-function in the light field. As an example of scalar point-functions, we may mention the *space illumination* (or space-density of radiant energy, which differs from the former by a constant factor).

5. The Illumination-distribution Solid

The most complete description of the light field at a given point is provided by the brightness-distribution solid. When this body is known, the illumination may be computed on an arbitrarily oriented element of surface at the point in question. Marking off the value of illumination along the normal to the illuminated element and rotating that element about P, so that its normal occupies all possible positions in space, we obtain the illumination-distribution surface (solid). The section of the surface by a plane passing through P gives an illumination-distribution curve in polar coordinates. When the field possesses axial symmetry, the illumination-distribution solid is characterized by a single curve. The concepts of the illumination-distribution solid and the brightness-distribution solid were introduced by L. Weber,[8] but his paper was unknown to illumination engineers until 1928, when Weber's concepts were revived by Lingenfelser.[9]

Let us give a few examples of the illumination-distribution solid for various simple cases of lighting. The outer normal to the illuminated surface will be used, as is customary in illumination engineering, though for simplicity the inner normal would be preferable.

(*a*) *A single point' source.* The illumination of an element at P, (Fig. 6) with its normal in the direction PI is

$$E = D_n = I/r^2 \tag{24}$$

[8] Leonhard Weber, Intensitätsmessungen des diffuses Tageslichtes. Ann. d. Phys., **26**, 1885, p. 374.

[9] H. Lingenfelser. Über den diffusen Anteil der Beleuchtung und ihre Schattigkeit. Licht und Lampe, 1928, No. 9, p. 313; Zur Messung und Beurteilung der räumlichen Beleuchtung. Licht und Lampe, 1930, No. 12, p. 619. (Also in Technisch-wissenschaftliche Abhandlungen aus dem Osram-Konzern, **2**, 1931, p. 143.)

When the element is inclined by the angle θ, the illumination is

$$E = D_n \cos \theta \tag{25}$$

The geometrical locus of the ends of the arrows representing the illumination E is a sphere with the diameter D_n. The illumination-distribution solid is a sphere which is tangent to a plane passing through P, the plane being normal to the incident light. The radius vectors on the other side of this plane are equal to zero. Eq. (25) holds only for those values of θ for which $\cos \theta > 0$ (for example, in the region $0 < \theta < \pi/2$). For values of θ for which $\cos \theta < 0$ (for example in the region $\pi/2 < \theta < \frac{3\pi}{2}$) the illumination is zero. Graphically, the dependence

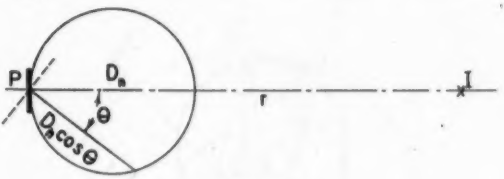


FIG. 6

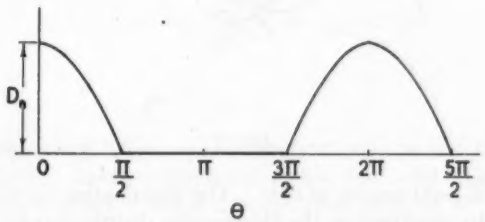


FIG. 7

of E on the angle θ is represented in rectangular coordinates by a periodic curve (Fig. 7), the illumination of the opposite side of the element being represented by a similar curve displaced by the angle π. A Fourier-series development of this function is

$$E = D_n \left[\frac{\cos \theta}{2} + \frac{1}{\pi} - \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k \cos 2k\theta}{(2k)^2 - 1} \right] \tag{26}$$

The expression in brackets is equal to zero for those values of θ for which $\cos \theta < 0$ and is equal to $\cos \theta$ if $\cos \theta > 0$. This law of variation

of illumination, because it is valid for all values of θ replaces the cosine law in the study of illumination-distribution solids.

The cosine law holds also for the difference of illumination of the two sides of a surface, that is, for the flux density. Negative illumination is absurd, but when we obtain a negative value for the flux density it means that the illumination is higher on the side opposite to the one upon which the normal has been constructed. The sinusoidal character of the variation of the difference of illumination in the case of a single point source insures the existence of the same law for any number of sources. This circumstance results in extraordinary simplicity.

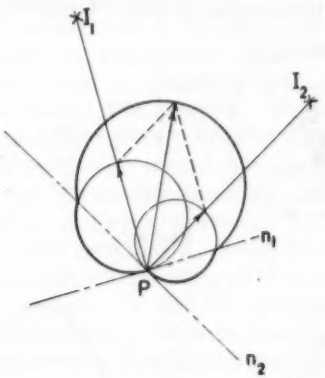


FIG. 8

(*b*) *Several point sources of light.* The illumination-distribution solid is obtained by constructing the illumination distribution for each of the sources separately, followed by an addition of the radius vectors. An example of two light sources is shown in Fig. 8. The illumination-distribution solid is bounded by three spherical surfaces. We recommend that the reader study in detail this important case, which we shall use in a number of further proofs.

(*c*) *The luminous hemisphere.* The distribution of illumination is determined at the point P, which is at the center of a uniformly luminous hemisphere of brightness B (Fig. 9). If the illuminated element is contained in the base-plane of the hemisphere, the illumination is

$$E_o = \pi B. \tag{27}$$

If the element is inclined by the angle ϕ, the illumination is

$$E = E_0 \frac{1 + \cos \phi}{2} = E_0 \cos^2 \phi/2 \qquad (28)$$

The values $\phi < \pi/2$ correspond to the upper side of the illuminated element, the values $\phi > \pi/2$ to the lower. This formula was derived by Lambert for the illumination of a space shadowed by an infinitely long wall and lighted by diffused light from the sky.

The illumination-distribution curve is given in this case by a cardioid. The illumination-distribution solid is obtained by rotating the cardioid about its axis. An identical illumination distribution would be produced by a uniformly bright plane of infinite extent.

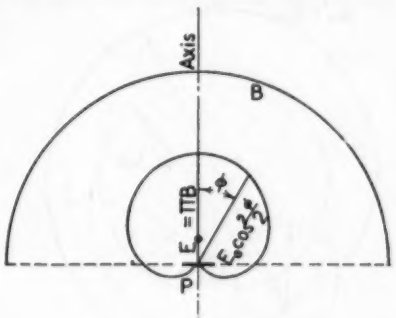


FIG. 9

The corresponding brightness-distribution solid is obtained in the following manner: In the directions within the solid angle 2π, the brightness has a constant value different from zero; in the other half the brightness is zero. Thus the brightness-distribution solid is represented by a hemisphere with its center at the point P.

After having discussed in his paper the case of the point-source and that of the luminous hemisphere, Weber proceeds to consider the illumination-distribution solid for the natural illumination from the open sky. In first approximation, the sun may be considered as a point source and the sky as a uniformly bright hemisphere. The section of the illumination-distribution solid by a vertical plane passing through the sun and through the location of the observer is obtained, as shown in Fig. 10, by addition of the radius vectors drawn to the circle (resulting from the sun) and those drawn to the cardioid (resulting from the sky).

The distribution of the illumination resulting from the sky and sunlight is shown by the heavy line.

(*d*) *The luminous sphere.* If the brightness of the inner surface of the sphere is B, the illumination-distribution surface at any point within the sphere is again a sphere with a radius πB and a center at the point considered. This may serve as an illustration of completely diffused light, where the brightness in all directions is the same. The brightness-distribution solid is a sphere with its center at P. This case is the opposite of the first case of parallel light, in which the brightness is zero in all directions but one. In both cases the illumination-distribution solid is a sphere, but in one case P is placed at the center of this sphere, and in the other case it is found on the surface of the sphere.

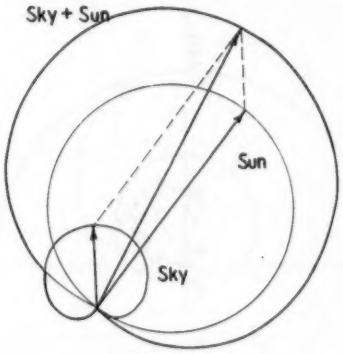


FIG. 10

Experiments on the determination of illumination-distribution solids for various important arrangements of artificial lighting were carried out by Lingenfelser.[9] He measured the illumination at the center and in the corner of an experimental room with direct, semi-indirect, and indirect lighting, with light and dark room walls, and with the receiving surface tilted at various angles. For the determination of the lighting conditions, he suspended two small white balls in the room. The shadows formed by the spheres on the walls and on the floor, compared with the shadows formed on the spheres themselves, showed the differences between the "hard" direct lighting and the "soft" indirect lighting. In Fig. 11 the solid line indicates the axial cross-section of the illumination-distribution solid at the center of the room for indirect lighting. For this point the illumination-distribution solid was found

to be a body of revolution, to within an accuracy of ten or fifteen per cent. For a point in the corner of the room, this symmetry does not exist.

Lingenfelser remarks in his article, quite correctly, that the main purpose in constructing illumination-distribution solids is not the quantitative evaluation of illumination, but the consideration of qualitative properties of lighting (degree of diffusion, production of shadows, etc.). Space forbids further consideration of Lingenfelser's analysis of illumination-distribution solids, but the method is presented briefly in the treatise on illuminating engineering by Sirotinsky and Fedorov.[10]

The illumination-distribution solid does not completely define the lighting conditions at a given point of the field, as was assumed by Weber. Knowing only the illumination-distribution solid, we cannot

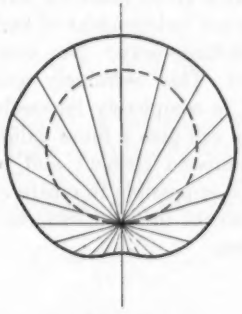


FIG. 11

compute the distribution of illumination on a surface of a real object because the various elements of the surface of the object are shading each other unless the body is convex.

If the effect of shadows of the object is to be obtained, the brightness-distribution solid must also be known. The distribution of brightness allows us to determine uniquely the distribution of illumination. The inverse problem of finding the brightness distribution, when the illumination-distribution solid is given, has no unique solution. Usually, the same illumination-distribution solid may be obtained for quite different brightness-distribution solids. This can be shown in a simple way in two-dimensional photometry, where brightness is measured as

[10] L. I. Sirotinsky and B. F. Fedorov, Principles of Electric Lighting. Energoizdat, 1923.

an intensity per unit length of the source, and illumination is inversely proportional to the first power of the distance. It may happen also that a distribution of sources cannot be found to satisfy arbitrary conditions of illumination. A question of fundamental interest and importance is this: May the illumination-distribution solid have an arbitrary form; may one choose an arbitrary illumination-distribution solid, assuming that for the point in question a corresponding brightness distribution can be found? Or the question may be formulated in the following manner: May one specify independent values of illumination upon the various inclined planes of a working surface, and assume that one can always find the lighting conditions which will result in this illumination distribution? The illuminating engineers answer this question in the affirmative. However, this is not quite right. The values of illumination at a given point for various orientations of the illuminated element are not independent of each other but are related, as we shall see, in a definite way. Not every kind of illumination distribution may exist! This seemingly unexpected conclusion is obvious *a priori*, for it is completely impossible to create at a given point such conditions as will give a finite value of illumination for one orientation of the illuminated element and zero illumination for all other orientations. It is impossible to create a horizontal illumination without illuminating also the inclined surfaces. We shall return later to this important problem.

CHAPTER III

Space Illumination

1. The Average Spherical Illumination

The analogy between the illumination-distribution solid and the curves of illumination distribution at a point in the light field, on the one hand, and the solid and curves of candlepower distribution for a point source, on the other hand, is obvious. One may also introduce the concept of *average illumination at a point*—a concept analogous to that of average intensity in a solid angle ω. The **average illumination** at P is defined as

$$E_\omega = \frac{1}{\omega} \int_\omega E \, d\omega \tag{29}$$

where E is the illumination of an element that is normal to the axis of the infinitesimal solid angle $d\omega$ having its apex at P. The integration is performed for all directions contained within the solid angle ω.

The **average spherical intensity** of a source may be defined as

$$I_{4\pi} = \frac{1}{4\pi} \int_{(4\pi)} I \, d\omega \tag{30}$$

where I is the intensity in the direction of the elementary solid angle $d\omega$ with its apex at the source, and where the integration is performed over all directions of space. Analogous to this, one may introduce the concept of **average spherical illumination** at P,

$$E_{4\pi} = \frac{1}{4\pi} \int_{(4\pi)} E \, d\omega \tag{31}$$

where $d\omega$ is the infinitesimal solid angle with its apex at P, and E is the illumination of an element normal to the axis of the solid angle.

The average spherical illumination is a fundamental concept for the light field. Indeed, the condition of the field at a point must be characterized by a quantity which has a single value at each point in space. The average spherical illumination satisfies this condition, whereas the common concept of illumination depends not only on position in space but also on the orientation of a plane element at P.

One may introduce also a concept of **average hemispherical illumination** at P:

$$E_{2\pi} = \frac{1}{2\pi} \int_{(2\pi)} E \, d\omega \tag{32}$$

or, for example, of the average illumination of a given cross section of the illumination-distribution solid (analogous to average horizontal candlepower of a source).

One may visualize in the following manner the concept of the illumination-distribution solid and of quantities which are defined by it. Let us describe a sphere about P, the radius being so small that the variation of illumination may be neglected from one point within the sphere to another. The surface of the sphere contains elements oriented in all possible ways in space. The illumination distribution at the point P may be considered as the distribution of illumination on the surface of the sphere.

This leads to a lucid interpretation of average values of illumination for a given ensemble of orientations of the illuminated element. Thus, for example, the average spherical illumination at a point P of the light field may be defined as the average illumination of the outer surface of a sphere of infinitesimal radius with its center at P. Let us denote by dF the luminous flux incident on an element of spherical surface,

$$ds = r^2 \, d\omega$$

where r is the radius of the sphere and $d\omega$ is the solid angle defined by the point P and the element ds. If E is the illumination of this element,

$$dF = E \, ds = r^2 E \, d\omega$$

and the flux incident on the entire surface of the sphere is

$$F = r^2 \int_{(4\pi)} E \, d\omega \tag{33}$$

Therefore the average illumination of the surface is equal to $E_{4\pi}$.

This proof may be generalized by stating that the average illumination at a point P in the light field, and within the boundaries of a definite solid angle, is the same as the average illumination of a part cut out by this solid angle from a sphere of infinitesimal radius with center at P. Thus we may introduce into illuminating engineering a new quantity, namely, the average hemispherical illumination, which was invented

by the author to parallel the concept of the average spherical illumination.[11] As indicated by its name, this quantity represents the average of the values at the given point for an element, the normal to which assumes all possible directions within the solid angle 2π. Thus the average hemispherical illumination is the limit of the average outside illumination on a hemisphere, with its center at the point P, when the radius approaches zero. It is obvious that the average spherical illumination depends only on the position of the point in the light field, whereas the average hemispherical illumination is also a function of the direction of the normal to the base plane of the hemisphere. Usually, however, the average illumination of the upper hemisphere with horizontal plane base is used.

When the illumination-distribution solid is known, the average values of the illumination within a given solid angle may be found by computation. For example, when the illumination-distribution solid possesses axial symmetry, it is completely described by the illumination-distribution curve obtained by a meridional section of it. The illumination at a given point depends only on the angle ϕ between the normal of the illuminated element and the axis of symmetry. When the equation of the curve of distribution of $E = E(\phi)$ is given in polar coordinates, the average spherical illumination may be defined as

$$E_{4\pi} \doteq \frac{1}{2}\int_0^{\pi} E(\phi) \sin \phi \, d\phi \tag{34}$$

For a symmetrical source, the average spherical intensity is obtained from the intensities in different directions by a similar formula. Analytical and graphical methods may be used to evaluate the integral, as in the computation of total flux from a Rousseau diagram.

2. Methods of Measurement

Instruments may be built for the direct determination of average illumination at a given point. For this purpose one may use a receptor of radiant energy which has a light-sensitive surface in the shape of a sphere or a hemisphere. A barrier-layer photocell of spherical shape suggests itself. The realization of such a photoelement, possessing absolutely uniform properties over its entire surface, is difficult; but one may use a polyhedron as is sometimes done in constructing apparatus for the measurement of luminous flux. This polyhedron may

[11] A. Gershun, Characteristics of Conditions of Illumination, Trans. Optical Institute, Leningrad, **6**, No. 59, 1931.

consist of similar, plane, barrier-layer cells. Such a quasi-spherical photoelement has been constructed by the research laboratory of Tungsram Company.[12] It consists of twelve selenium cells which form a regular icosahedron. The elements are connected in parallel, and the indication of the associated electrical instrument is proportional to the sum of the photoelectric currents produced by the individual cells.

Apparatus for measurement of average spherical and hemispherical illumination may be constructed by using a somewhat different principle. Let us place in the light field a small hollow sphere made of a diffusing material which follows Lambert's law for transmitted light as well as for reflected light. The illumination of the inner surface of the diffusing envelope will be uniform over the entire surface, and this illumination will differ only by a constant factor from the average illumination of the outer surface of the sphere. The first statement follows directly from the known photometric property of the sphere. To prove the second statement, let us determine the illumination E_{in} of the inner surface of the sphere. Let τ and ρ be the transmission factor and reflection factor of the spherical shell, and let s be the surface area of the entire sphere. Denoting by $E_{4\pi}$ the average spherical illumination at a given point, we have a flux

$$F = E_{4\pi} s$$

on the outside of the sphere. Part of this flux

$$\tau F = \tau E_{4\pi} s$$

enters the sphere and is distributed over the entire inner surface. Multiple reflections must be considered, but the balance for the fluxes is given by the expression

$$\underbrace{E_{in} s}_{\text{on inner surface}} = \underbrace{\tau E_{4\pi} s}_{\text{from outside}} + \underbrace{\rho E_{in} s}_{\text{reflected from within}} \tag{35}$$

Thus,

$$E_{in} = \frac{\tau}{1 - \rho} E_{4\pi} \tag{36}$$

Therefore the illumination within differs only by a constant factor from the average illumination on the outside of the sphere. Knowing this factor, and having measured the illumination E_{in} at any point of

[12] Über eine Sperrschichtphotozelle zur Messung der Raumhelligkeit. (Mitteilung aus dem Tungsram-Forschungslaboratorium). Das Licht, 4, 1934, p. 155.

the inner surface of the sphere, we may determine the value of average spherical illumination. If absorption of light is neglected,

$$E_{in} = E_{4\pi} \tag{36a}$$

Thus in the determination of average spherical illumination one may use a diffusing sphere as has been done in the measurement of average spherical candlepower; but in the former case one measures the light transmitted through the sphere rather than the light reflected from it. A number of simple photometric devices can be used for the measurement of average spherical illumination, or more exactly speaking, of space-illumination, which differs from the former by a constant factor and which will be discussed later. The diffusing shell of the sphere is made of opal glass. One should remember, however, that opal glass is not the ideal diffusing material of which the foregoing imaginary sphere was made. Opal glass does not follow Lambert's law, and possesses also selective properties. Moreover, any arrangement for the measurement of illumination within the sphere necessarily blocks off part of the total surface. All these factors, together with the finite dimensions of the sphere, impose a limit on the possible accuracy of the measurement of average spherical illumination.

For measurement of illumination on the inner surface of the sphere, one may use objective methods as well as visual methods. For objective measurements one uses a receptor (photoelement, photographic plate), the light-sensitive surface of which constitutes a small part of the inner surface of the spherical probe. A simple photometer containing a barrier-layer photocell has been constructed on this principle by Lange.[13]

Arrangements for measuring average spherical illumination may be combined with any visual photometer* used in measuring brightness. The neck of a spherical flask made of opal glass (the author used envelopes of low-wattage opal-glass lamps) is attached to a metal tube (Fig. 12) which leads to the photometric screen. The end of the tube that terminates at the spherical surface is closed by a piece of opal glass. On this opal glass window falls only the light diffusing from the flask.

[13] B. Lange, Über die Photometrische Anwendbarkeit der Halbleiterphotozellen. Zs. für Instrumentenkunde, **53**, 1933, p. 380.

* Laboratory photometers, as well as most illumination meters, such as the one designed by the State Optical Institute and used extensively in the U. S. S. R. Attachments for use in measuring space illumination are manufactured by Franz Schmidt & Haensch according to the design of Arndt (Licht u. Lampe, **7**, 1928, p. 247).

The illumination of the glass, and therefore its brightness, is proportional to the average illumination of the sphere. When the sphere is sufficiently small, this illumination may be assumed to be proportional to the average spherical illumination at the center of the sphere. In this manner the measurement of average spherical illumination is reduced to a single measurement of brightness.

The same apparatus may be used not only for the measurement of $E_{4\pi}$ but also for the measurement of such magnitudes as E_ω. For the latter purpose one must enclose the measuring sphere in an opaque hood, leaving open only part of its surface. The whole inner surface of the sphere will be uniformly illuminated as previously, and this illumination will be proportional to the average outside illumination

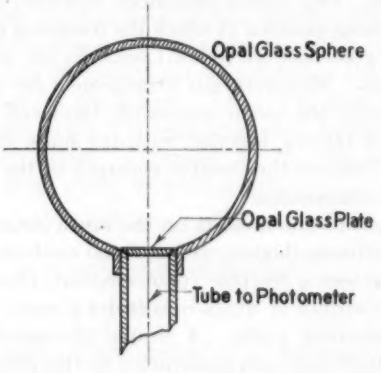


FIG. 12

of the unshielded part of the measuring sphere. Several such designs have been made by the author[11] in collaboration with V. A. Zelenkov and D. N. Lazareff. One of these attachments intended for measurement of average hemispherical illumination $E_{2\pi}$ is indicated in Fig. 13. A hemisphere of opal glass is attached to a cylindrical box, the height of which is equal to the radius of the hemisphere. The box is blackened inside. At the center of the bottom of the box there is an orifice closed by matt opal-glass plate. This window represents, let us say, an element of the sphere, the other part of which is the light-diffusing hemisphere. The observer looks through the tube and measures the brightness of the opal-glass plate. The reading of the photometer corresponding to the photometric balance is then multiplied by a con-

stant which corresponds to the particular attachment, and which had been previously determined by calibration. For the solution of problems of illuminating engineering it may be useful to introduce fractional magnitudes such as $E_{3\pi/2}$, E_{π}, $E_{\pi/2}$, as was done by D. N. Lazareff. For this purpose the opal-glass hemisphere is screened by various attachments.

3. Methods of Computing Space Illumination

Let us consider now in greater detail the newly introduced fundamental magnitude called the average spherical illumination. One can give the following interesting definition of this magnitude from which will follow directly the method of its computation if the brightness-

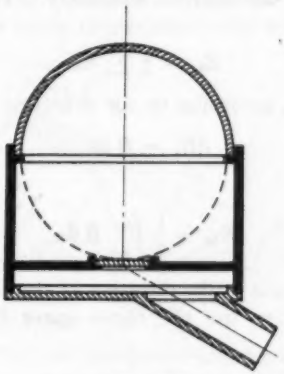


FIG. 13

distribution solid is known. Let us subdivide the entirety of all rays passing through a given point into a set of elementary solid angles $d\omega$. Let us denote by dD_n the normal illumination* produced by the rays that are contained within $d\omega$. The part of the flux received by the outer surface of a small sphere of radius r is

$$\pi r^2 \, dE_n$$

The entire flux received by the sphere from all directions of space is

$$F = \pi r^2 \int_{(4\pi)} dE_n \tag{37}$$

* [By "normal illumination" is meant the value of illumination when the light rays are perpendicular to the illuminated surface.] Trans.

The average spherical illumination for the point in question is

$$E_{4\pi} = \frac{F}{4\pi r^2} = \frac{1}{4}\int_{(4\pi)} dE_n \tag{38}$$

In this manner the average spherical illumination may be defined as a quarter of the sum of all elementary normal illuminations.

When the illumination is produced by a set of point sources,

$$E_n = \frac{I}{r^2} \tag{39}$$

where I is the intensity of the source in the direction considered, r is the distance from the point to the source, and E_n is the normal illumination* produced by each of the sources separately. The average spherical illumination is

$$E_{4\pi} = \tfrac{1}{4}\sum E_n \tag{40}$$

In general we have, according to our definition of brightness,

$$dE_n = B\,d\omega \tag{41}$$

and therefore,

$$E_{4\pi} = \frac{1}{4}\int_{(4\pi)} B\,d\omega \tag{42}$$

where B is the brightness in the direction of the elementary solid angle $d\omega$. It is customary to use the name **space illumination**[14] for the quantity

$$E_0 = 4E_{4\pi} = \int_{(4\pi)} dE_n = \int_{(4\pi)} B\,d\omega \tag{43}$$

Equation (43) allows us to determine the value of average spherical illumination when the brightness-distribution solid is known. The average hemispherical illumination $E_{2\pi}$ is

$$E_{2\pi} = E_{4\pi} + \frac{E_h - E'_h}{4} \tag{44}$$

as has been shown by the author.

[14] Introduced by Leonhard Weber, Die Albedo des Luftplanktons. Ann. d. Phys., **61**, 1916, p. 427. [The concept does not seem to have been used in the United States and no English name has been established for it. Russian, *prostranstvennaya osveshchennost*; German, *Raumbeleuchtung*; French, *éclairement spatial*.] Trans.

$E_{4\pi}$ is the average spherical illumination, while E_h and E_h' are the values of illumination of the two sides of the base of the hemisphere. If the upper hemisphere is considered, these values refer to the horizontal illumination from above and from below.

4. Space Density of Light*

Because light is propagated with a finite velocity, we find in each volume of the light field at any instant of time a definite quantity of energy and a definite number of light particles or *photons*. Therefore one may introduce a concept of *volume density of radiation*, as was done by Planck. Let us define the limit of the ratio of the quantity of light to the volume containing it, as this volume approaches zero, as the **space density of light** at the point in question. The space density of light differs from the space illumination only by a constant factor, a

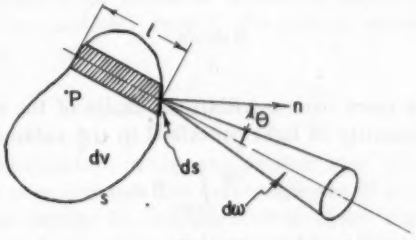


FIG. 14

constant of the medium, the medium being assumed to be isotropic.[11] Let us compute u, the space density of light, at the point P of the light field (Fig. 14). We assume for a moment that the light is removed from all the space with the exception of the elementary volume dv containing the point P; then we compute the energy which flows out of this volume through its surface s.[15]

The luminous flux through the element of the surface ds in the direc-

* ["Light energy" in the original. We are using the word "light" to mean *radiant energy evaluated with respect to the standard visibility function*. We use the photometric rather than the radiometric name in accordance with the policy stated in Chap. II. Throughout his book, Professor Gershun wishes to emphasize the intimate connection between illuminating engineering and radiation engineering and the fact that the same principles (conservation of energy, etc.) can be applied in both fields.] Trans.

[15] A. F. Joffe, Thermodynamics of Radiant Energy. Chapter 5 of the Treatise on Physics by O. D. Chwolson. (In Russian.)

tions contained within an elementary cone $d\omega$, the axis of which makes an angle θ with a normal n to the element ds, is equal to

$$B \cos \theta \, ds \, d\omega$$

This light will flow during the time l/c, where l is the distance along the light ray from the element ds to the opposite wall of the volume dv, and c is the velocity of propagation. Thus the beam will carry through ds the light equal to

$$\frac{Bl}{c} \cos \theta \, ds \, d\omega$$

The quantity $l \cos \theta \, ds$ represents the shaded element of volume shown in the figure. Integrating over the entire surface, we obtain the quantity of light that has emerged from dv in directions contained within $d\omega$, or

$$\frac{B \, d\omega \, dv}{c}$$

Integrating once more over ω within the limits of the solid angle 4π, we obtain the quantity of light contained in the volume dv, or

$$dQ = \frac{dv}{c} \int_{(4\pi)} B \, d\omega \tag{45}$$

Thus the space density of light is

$$u = \frac{dQ}{dv} = \frac{1}{c} \int_{(4\pi)} B \, d\omega = \frac{E_0}{c} \tag{46}$$

which differs by a constant factor from the space illumination. In the accepted system of units of illumination engineering, the space density of light is measured in *lumen sec* cm^{-3}.

5. The Concept of Space Illumination

As has been shown, the average spherical illumination $E_{4\pi}$, the space illumination E_0, and the space density of light u are equivalent concepts, and inasmuch as

$$E_0 = 4E_{4\pi} = cu \tag{47}$$

one need consider only one of these quantities. The space density of light is not a convenient quantity to deal with in engineering, one of the reasons being that its values are expressed by very small numbers

in the accepted system of units and the concept is not easily visualized by an engineer, who customarily deals with power and with luminous flux.

The average spherical illumination is expressed in units of illumination, the same units being used also for space illumination. According to Arndt the unit of space illumination should be called "Raumlux" to distinguish it from the ordinary lux. To obtain the number of "space lux" at a point, one must add the values of normal illumination (expressed in lux) which are caused by all the light sources. Space illumination represents a fundamental quantity characterizing the light field at a point, and this concept must find a wide application in illuminating engineering practice. Space illumination characterizes the general "density of light" at a given point in a room, independent of the direction from which the light is coming; in other words, it characterizes the average of all values of brightness taken in all directions from the point in question. Indeed, the average spherical brightness $B_{4\pi}$ is defined as

$$B_{4\pi} = \frac{1}{4\pi} \int_{(4\pi)} B \, d\omega = \frac{E_0}{4\pi} \tag{48}$$

Thus space illumination defines the average level of brightness seen from the point in question. The value of space illumination at each point in a room lighted by luminaires that throw the light in comparatively narrow beams is conveniently divided into two parts—the first produced by direct light from the luminaires, and the second produced by light reflected from surfaces in the room. The ratio of these two parts characterizes approximately the degree of direct lighting and the depth of the shadows. The value of the second part gives the average level of brightness of the enclosing surfaces and of the surfaces of objects contained within the room. Thus it determines the adaptation level of the eye and defines the visual evaluation of all brightness. The actual evaluation of the adequacy and uniformity of lighting as it appears to the eye does not correspond to the evaluation of lighting based on the illumination distribution on a fictitious working surface, but is more closely characterized by the space illumination.

The transition from standardization according to horizontal illumination to standardization according to space illumination is completely analogous to the change from candlepower rating to lumen rating—a change that is universally accepted in the specification of lamps. The transition to space illumination is desirable in the natural

lighting of buildings. At the present time the adequacy of natural lighting is judged by the "coefficient of natural lighting," which is equal to the ratio of horizontal illumination within the room to the corresponding value of outside illumination. Such a choice of horizontal illumination as a measure produces an unjustifiable discrepancy in the comparison of lighting from windows and from luminaires. Much larger values are obtained for coefficients of lighting in the case of lighting from above, than in the case of lighting from the side. This is due to the fact that these coefficients are arbitrarily taken relative to the horizontal plane. The side light from the windows falls on the horizontal surfaces at large angles of incidence and thus produces low values of illumination, though it may well fill the entire room with light. This consideration suggests the desirability of using the concept of space illumination for the specification of daylighting, as has been pointed out in a number of foreign articles.* One must note especially the articles of Arndt, which are summarized in his book, *Raumbeleuchtungstechnik*.

Space illumination is an average of the values of illumination for all possible orientations of the illuminated surface. One usually looks upon objects from above, however, and thus it seems reasonable in many cases to judge the illumination by the value of the upper hemispherical illumination. In the choice of the characteristics of illumination one also has to consider the micro-relief of the surface in question, together with its macro-orientation and the space orientation of sepa-

* Wilhelm Arndt, Raumhelligkeit als neuer Grundbegriff der Beleuchtungstechnik. Licht u. Lampe, 1928, No. 7, p. 247.

Wilhelm Arndt, Beleuchtungsstärke oder Raumhelligkeit. Licht u. Lampe, 1928, No. 23, p. 833.

Wilhelm Arndt, Neue Grundzüge der Beleuchtungstechnik. Licht u. Lampe, 1930, No. 10, p. 537.

Wilhelm Arndt, Über den Stand der Arbeiten, Beleuchtung räumlich zu bewerfen. I. I. C. Proc., 1931, p. 197.

Wilhelm Arndt, Raumbeleuchtungstechnik, Berlin, 1931.

* A. Dresler and W. Arndt, Beleuchtungswertung mit Hilfe der Raumhelligkeit. Licht u. Lampe, 1930, No. 20, p. 997; No. 22, p. 1092; No. 23, p. 1142.

A. Dresler, Entwurf und Beurteilung von Beleuchtungsanlagen auf räumlicher Grundlage. Dissertation, Tech. Hochschule, Berlin, 1930.

H. Hellmann, Die Bewertung von Beleuchtungsanlagen auf messtechnischer Grundlage. Dissertation, Berlin, 1930.

Jean Dourgnon, Définition et calcul des grandeurs caracteristiques de l'éclairage d'un éspace clos. R. G. E., **33**, 1933, p. 579.

rate elements forming its structure.[16] In evaluating the illumination of an aeroplane landing field covered with grass, for instance, one must consider not only the horizontal illumination of the earth but also the vertical illumination of the blades of grass.

[16] Gershun and Lazareff, Light Engineering (Russian periodical), No. 4, 1935, p. 1.

CHAPTER IV

The Light Vector

1. Proof of Existence

The luminous flux through a surface may be considered as the flux of a vector point-function, the projection of this vector on a given direction being proportional to the difference in illumination of the two sides of a plane element which is normal to that direction.

Let us consider a light field in an arbitrary medium. This medium may diffuse the light that is passing through it, or it may emit and absorb light. Enclose a point of the light field by a small surface. Reduction in the size of the surface reduces the enclosed volume; and if the surface remains geometrically similar to its previous shape, then the luminous flux through the surface will decrease in proportion to the area of the surface. But the radiation absorbed or emitted within this surface will decrease much more rapidly, i.e. in proportion to the volume enclosed by the surface. When the linear dimensions are decreased to $1/n$, the surface will decrease to $1/n^2$ and the volume to $1/n^3$. Thus a sufficiently small volume can always be found to justify the assumption that all the flux entering the region is leaving the region.

The proof of the existence of a vector point-function defining the flux of radiant energy is completely analogous to the proof used in the analytic theory of heat.[17] The method was applied to photometry by Boldyreff.[18] Consider a point P with coordinates x, y, z, measured with respect to a rectangular system of coordinates X, Y, Z, (Fig. 15). Form the elementary tetrahedron having three planes meeting at the point P, these planes being parallel to the planes formed by the coordinate axes. The distances dx, dy, dz, are assumed to be sufficiently small so that the illumination is uniform over any one of the four sides of the tetrahedron. We assume also that absorption and emission of light *within* the elementary volume may be neglected.

[17] G. Kirchhoff, Vorlesungen über die Theorie der Wärme. Leipzig, 1894.
I. Boussinesq, Théorie analytique de la chaleur. Paris, 1901, (Vol. I, §§51-54).
H. Helmholtz, Vorlesungen über die Theorie der Wärme. Leipzig, 1903 (Vol. 6, Part 2, Chapter I).

[18] N. G. Boldyreff, The Light Field in Diffusing Media. Trans. Optical Institute, Leningrad, 6, No. 59, 1931.

Let us denote by E_x, E_y, E_z, the illumination of the three sides that meet at the point P. The illumination of the fourth side is E_n, where n denotes the normal to that side. To distinguish between the two sides of each of the four surfaces, we shall add an index plus when we consider the side of the surface that is seen from the origin of coordinates, and an index minus for the opposite case. Thus the illumination of the sides of the tetrahedron from outside will be E_{+x}, E_{+y}, E_{+z}, E_{-n}. Correspondingly, we shall denote by E_{-x}, E_{-y}, E_{-z}, E_{+n}, the illuminations from within. Let us compute the fluxes entering and leaving the tetrahedron. If $d\sigma_x$, $d\sigma_y$, $d\sigma_z$, $d\sigma_n$ are the areas of the four sides, the entering flux is

$$dF_{ent} = E_{+x}\,d\sigma_x + E_{+y}\,d\sigma_y + E_{+z}\,d\sigma_z + E_{-n}\,d\sigma_n \tag{49}$$

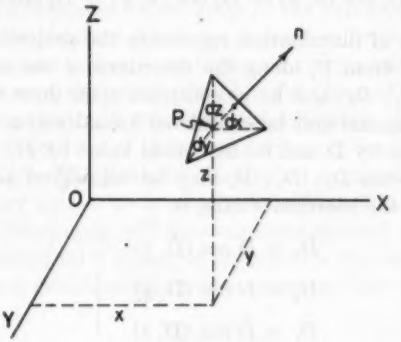


FIG. 15

and the flux that leaves is

$$dF_{leave} = E_{-x}\,d\sigma_x + E_{-y}\,d\sigma_y + E_{-z}\,d\sigma_z + E_{+n}\,d\sigma_n \tag{50}$$

According to our assumptions, dF_{ent} equals dF_{leave}. Therefore,

$$\begin{aligned}(E_{+n} - E_{-n})\,d\sigma_n = (E_{+x} - E_{-x})\,d\sigma_x + (E_{+y} - E_{-y})\,d\sigma_y \\ + (E_{+z} - E_{-z})\,d\sigma_z .\end{aligned} \tag{51}$$

This equation states that the luminous flux through an arbitrarily oriented element of surface is equal to the algebraic sum of the fluxes through the projection of this element on three planes which are perpendicular to each other and which pass through the point in question. The orientation of these planes is arbitrary. The flux through a surface

is defined, as usual, by the difference of fluxes on the two sides of the surface. Dividing Equation (51) by $d\sigma_n$ and introducing the expression for the cosines of the angles between the normal n and the coordinate axes, we obtain

$$(E_{+n} + E_{-n}) = (E_{+x} - E_{-x}) \cos(n, x) + (E_{+y} - E_{-y}) \cos(n, y) + (E_{+z} - E_{-z}) \cos(n, z) \cdots . \tag{52}$$

Denoting the flux density through the surface (the difference of illumination) by D, we may write

$$D = E_+ - E_- .$$

Thus

$$D_n = D_x \cos(n, x) + D_y \cos(n, y) + D_z \cos(n, z). \tag{53}$$

The difference of illumination represents the projection of a vector. Let us mark off from P, along the directions of the coordinate axes, distances D_x, D_y, D_z, and let us construct upon these three a parallelopiped. The diagonal may be considered formally as a vector. Let us denote the vector by $\mathbf{D}$ and its numerical value by D.

The flux densities D_x, D_y, D_z may be considered as projections of the vector $\mathbf{D}$ on the coordinate axes, or

$$\left.\begin{aligned} D_x &= D \cos(\mathbf{D}, x) \\ D_y &= D \cos(\mathbf{D}, y) \\ D_z &= D \cos(\mathbf{D}, z) \end{aligned}\right\} \tag{54}$$

where

$$D = \sqrt{D_x^2 + D_y^2 + D_z^2} \tag{55}$$

is the absolute magnitude of the vector. The direction cosines of the vector are equal to

$$\frac{D_x}{D}, \quad \frac{D_y}{D}, \quad \frac{D_z}{D}$$

Substituting these expressions for D_x, D_y, D_z in Equation (53), and introducing the expression for the cosine of the angle between the vector $\mathbf{D}$ and the normal n, which is equal to

$$\cos(\mathbf{D}, n) = \cos(\mathbf{D}, x) \cos(n, x) + \cos(\mathbf{D}, y) \cos(n, y) + \cos(\mathbf{D}, z) \cos(n, z) \cdots \tag{56}$$

we obtain the following fundamental relationship:

$$D_n = D \cos (\mathbf{D}, n) \tag{57}$$

which is independent of the choice of coordinate axes.

Inasmuch as the normal n has been chosen arbitrarily, we may state the following: *At each point P of the light field there is a vector, which is independent of the choice of coordinate axes, and which possesses the property that its projection upon any direction is numerically equal to the difference of illumination of the two sides of a plane element placed at P and normal to that direction.* This vector and the space-illumination function represent fundamental functions of position in the light field. The space illumination is a scalar point function, while the vector is a vector point function. We shall call this vector the **light vector**.

The greatest difference of illumination of the two sides of an element at a given point is obtained when the element is perpendicular to the vector. This position corresponds to the maximum transfer of energy. When the illuminated element is placed in the plane of the vector, it is equally illuminated on both sides and the difference of illumination is zero. There is no transfer of energy and dynamic equilibrium obtains. At each point of the light field the illuminated element may be rotated about an arbitrary axis so as to obtain equal illuminations on the two sides. That is, the element will determine the plane of the light vector. Photometric equilibrium will not be destroyed if the element in such a position is rotated about an axis which coincides with the direction of the light vector.

2. Properties of the Light Vector

Let us consider a geometrical surface s in the light field (Fig. 16). By a geometrical surface is meant a surface that possesses no physical properties, and which therefore does not affect the light field. As a rule the surface s will be illuminated from both sides. We shall denote the two sides by subscripts 1 and 2. The illuminations of the two sides of an element ds are E_1 and E_2, and the light vector for the element ds is $\mathbf{D}$. The angle between $\mathbf{D}$ and n (the sense of the normal is indicated in Fig. 16 by an arrow) is denoted by θ. Then

$$E_1 - E_2 = D \cos \theta. \tag{58}$$

Multiplying the left and the right sides of this equation by ds, we obtain

$$E_1\, ds - E_2\, ds = D \cos \theta\, ds. \tag{59}$$

The expression on the left side of the equation represents the difference of the flux

$$dF_1 = E_1\,ds$$

on the side No. 1 of ds, and the flux

$$dF_2 = E_2\,ds$$

on the other side. Denoting this quantity by dF, we shall call it the luminous flux through the element ds.

Consider now the right side of Equation (59), representing the flux of the vector **D** through the surface element ds. Introducing the concept of a vectorial element of surface $d\mathbf{s}$, which is equal in magnitude to ds, and which is directed in the direction of the normal n,

$$dF = \mathbf{D}\cdot d\mathbf{s}.$$

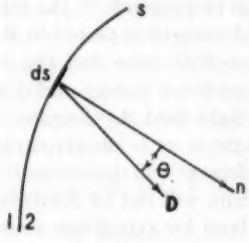


FIG. 16

Integrating this expression over the entire surface s,

$$F = \int_s \mathbf{D}\cdot d\mathbf{s} \qquad (60)$$

In the terminology of vector analysis the luminous flux through the surface is equal to the flux of the light vector through that surface.[19]

This theorem may be considered as the fundamental definition of the light vector. The existence in the light field of a vector point-function, the integral of which over a surface represents the quantity of energy passing through the surface, must be considered as the initial and most fundamental theorem of theoretical photometry. But today it is still not universally accepted—a state of affairs that may astonish

[19] A. Gershun, The Light Field Resulting from Luminous Surfaces of Uniform and Non-uniform Brightness. Trans. Opt. Inst., Leningrad, **4**, No. 38, 1928.

a physicist who is not acquainted with the literature on photometry. We may recall that in his study of the electromagnetic field of an elementary vibrator, Poynting showed in 1885 that the increase of energy in an arbitrary volume is measured by the flux of a vector through the surface enclosing this volume. The light vector, introduced in the present treatise, may be considered as a macroscopic average (in space and in time) of the values of the Poynting vector. Obviously the light vector characterizes the magnitude and direction of the light pressure on a small spherical body placed at the point in question.

The light vector has been defined above as a vector, the projection of which upon a given direction is equal to the difference of illumination of the two sides of an element that is normal to that direction. In other words, the projection of the light vector determines the flux density, which characterizes the energy balance and which is of great importance in the theory of the physical field. However, the illuminating engineer is not interested in energy flow through imaginary apertures but in the illumination of real bodies, especially in the illumination of their outer surfaces. From this standpoint, the engineer may doubt the soundness of introducing the new concept of the light vector. The author hopes that if the reader has a similar prejudice it will disappear as a result of the study of this book. Before we take up the matter of photometric computation, let us illustrate several applications of the vector method to show its usefulness in illuminating engineering.

As the first example let us consider the following problem. Two point sources of light of intensity I_1 and I_2 are at distances r_1 and r_2 from a point P (Fig. 17). What is the maximum value of the illumination at P, and how should the illuminated element be placed to obtain on it the maximum illumination? Without the concept of the light vector we should have to derive an expression for the illumination of the element (arbitrarily oriented) due to each of the sources, to write an expression for the sum of two such effects, and to find by means of differential calculus the position of the element, for which its total illumination has a maximum value. All this is not very difficult, but that it is sufficiently involved is indicated by the absence of the treatment of this problem (which seems to be a fundamental one) in the scientific literature of photometry.

Let us solve the problem by the vector method. We denote the two light vectors by $\mathbf{D}_1$ and $\mathbf{D}_2$. Obviously the directions of these vectors coincide with the directions of the rays arriving at P from the sources.

The absolute values of these vectors are

$$D_1 = \frac{I_1}{r_1^2}, \qquad D_2 = \frac{I_2}{r_2^2}$$

By means of the parallelogram rule the resultant light vector **D** is equal to

$$\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2 . \tag{61}$$

For all orientations for which both I_1 and I_2 are on the same side of the illuminated element, the illumination of the other side is zero; and thus the projection of the vector on the normal to the surface is equal

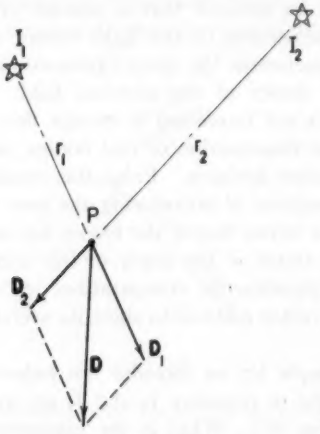


FIG. 17

to the illumination of the surface. The illumination resulting from each of the sources considered separately varies as the cosine of the angle between the vector $\mathbf{D}_1$ or $\mathbf{D}_2$ and the normal to the illuminated element. The sum (or difference, when both sides are illuminated) varies as the cosine of the angle between **D** and n. The total illumination has its maximum (which is equal to the length of the vector **D**) when n is collinear with **D**. Thus the problem is solved. Its solution consists of the construction of a parallelogram and the determination of the length of its diagonal, which may be performed either graphically or analytically. *The parallelogram of illuminations* may play in photometry a rôle similar to that played by the *parallelogram of forces* in mechanics.

The distribution of total illumination in this case is shown in Fig. 8 by the heavy line. The same figure shows the parallelogram of illuminations, but the vectors are marked off in directions opposite to those chosen in Fig. 17. The curve representing the distribution of illumination, resulting from each of the sources separately, is a circle. The total illumination is represented again by a circle (the heavy curve, with the exception of the part of it enclosed between n_1 and n_2; this latter part corresponds to those positions of the illuminated element where the sources are placed on both sides of it). The diameter of this circle is obtained as the vector sum of the diameters of the other two circles.

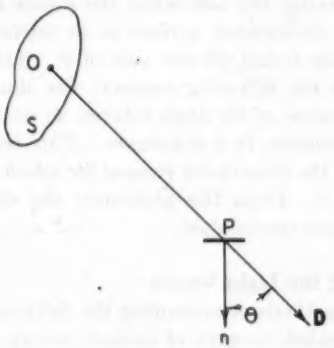


FIG. 18

As a second example, we shall consider the illumination resulting from large luminous surfaces. Let us determine the illumination at the point P (Fig. 18); that is, let us determine the illumination of an element s at the point P, as a function of orientation. The illumination E of the element s is directly proportional to cos θ, or

$$E = D \cos \theta.$$

Thus the computation of illumination is reduced to the computation of the light vector. The straight line OP may be called the light axis, and the point O, where the continuation of the vector intersects the surface S of the luminous body, may be called the *light center of the body S for the point P*. As a rule, the position of the light center varies as P moves in the light field.

If the illuminated element s is placed in such a manner that its plane

intersects the luminous body S and divides it into two parts, S_1 and S_2, the projection of the vector is equal to the difference of illumination of the two sides of the element. Let E_1 be the illumination on side 1 (which is produced by S_1) and let E_2 be the illumination of side 2 (produced by S_2). Then

$$E_1 - E_2 = D \cos \theta. \tag{58}$$

This relationship does not allow the evaluation of E_1 and E_2, though it enables the illumination on one side to be determined when the illumination of the other side is known.

The cosine law may be applied not only when illumination is produced by a point source but also when the source is an extended surface, provided the illuminated surface is so oriented that all of the source or sources are found on one side of it. Let us state this important theorem in the following manner: *The illumination of an element varies as the cosine of the angle between its normal and the normal for which the illumination is a maximum. This result is valid for all those orientations of the illuminated element for which all the light sources are on one side of it.* From this statement the eleventh theorem of Lambert[20] follows as a special case.

3. Measurement of the Light Vector

Let us consider methods of measuring the light vector. Suppose we have a plane, two-sided receptor of radiant energy, combined with an instrument that allows us to determine equality of illumination on the two sides of the receptor. The sensitivity of the receptor must of course be the same on both sides. Such an arrangement may be accomplished by means of a plane set of thermocouples, with junctions on both sides, connected to a galvanometer; or we may use a visual photometer and judge the photometric balance by the disappearance of the brightness difference. Or we may use an objective photometer, such as a two-sided photoelement, connected to a galvanometer so as to read the difference between the illuminations of the two sides of the photoelement. These receptors may be classed as the *differential type*, since they respond to the difference in the illuminations of the two sides.

The apparatus for the determination of the light vector must be constructed in such a manner that the plane of its receptor may be oriented arbitrarily in space. Rotating the receptor about two axes,

[20] Ostwald's Klassiker d. exakten Wissen. No. 31, 1892, p. 53; No. 33, p. 84.

we find two different positions of the receptor plane (s_1 and s_2, Fig. 19), for which there is a photometric balance. The light vector must be contained in the plane s_1 as well as in the plane s_2, and thus it is directed along the intersection of the two planes. The direction of the vector may be characterized by two angles (for example, by the longitude and the latitude) and provision to read these angles must be made in the construction of the apparatus. Placing the plane of the receptor in a position s, which is normal to the direction of the light vector, and determining the difference of illumination of the two sides of it, we find the absolute magnitude of the vector, which we shall express in units of illumination.

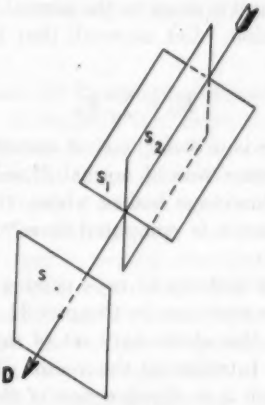


FIG. 19

One may also determine the light vector indirectly by measurement of the illumination on the two sides of the element for three different orientations of the element; i.e., by measurement of the components of the vector in the corresponding inclined system of coordinates. Most convenient is the choice of three mutually perpendicular directions. This may be interpreted as measurement of illumination on the six sides of a small cube, which is placed at the point in question. When the three components of the vector are determined in this way, one knows the magnitude and the direction of the vector.

The first apparatus for the determination of the light vector, called the *Vectorscope*, was constructed at the Optical Institute by M. M. Gurevitch. Another apparatus, of simpler and improved type, was

constructed by V. A. Zelenkoff (Illuminating Engineering Laboratory of the State Optical Institute) at the request of the author.

4. Computation of the Light Vector

We consider now the methods of computation of the light vector. For a light field produced by a set of point sources, the resultant light vector **D** at a point P is defined as the sum of the light vectors $\mathbf{D}_n$ of each of the sources, or

$$\mathbf{D} = \sum \mathbf{D}_n$$

Each of these component vectors $\mathbf{D}_n$ is directed along the ray from the source to P, and its length is equal to the normal illumination produced by the source in question. Let us recall that the space illumination E_0 has been defined as

$$E_0 = \sum E_n$$

The space illumination is a *scalar sum* of normal illuminations, while the light vector is a *vector sum* of normal illuminations. If the light field is produced by luminous bodies whose angular size cannot be neglected, the light vector is computed from the known brightness-distribution solid.

Let us subdivide the entirety of rays passing through a point into elementary bundles, as was done in Chapter II. Let dE_n be the illumination produced by this elementary set of rays on a plane perpendicular to their axis. Introducing the vector $d\mathbf{D}_n$, which is equal in length to dD_n and which is in the direction of the propagation of light along the axis of the elementary cone, we obtain

$$\mathbf{D} = \int_{(4\pi)} d\mathbf{D}_n$$

where (4π) indicates that the integration is performed over all directions of space. Denoting by $d\omega$ the solid angle of the elementary cone, we obtain

$$dE_n = B\, d\omega.$$

Let us introduce a vector $d\boldsymbol{\omega}$, which is equal in magnitude to the angle $d\omega$ and which is directed along the axis of the elementary cone. Using this geometric concept, we obtain

$$d\mathbf{D}_n = B\, d\boldsymbol{\omega}$$

and

$$\mathbf{D} = \int_{(4\pi)} B \, d\omega \tag{62}$$

Let us recall that the space illumination is defined by the relationship

$$E_0 = \int_{(4\pi)} B \, d\omega \tag{43}$$

Thus one may state that the space illumination represents a *scalar integral* of brightness, taken over the solid angle 4π, and that the light vector is a *vector integral* of brightness over the same solid angle. These two quantities are the fundamental functions of position in the light field.

5. The Light Vector and the Illumination-distribution Solid

Let us illustrate the relationship that exists between the light vector and the illumination-distribution solid. This will also give an answer to the question, whether the illumination-distribution solid may be of any arbitrary shape. The flux density represents the projection of the light vector. From this follows that the flux-density distribution solid (the entirety of radius vectors representing the differences of illumination on the two sides of an element which is perpendicular to them) is always a sphere; and the point P, for which this body was constructed, is found on the surface of the sphere. The diameter of the sphere gives the direction and magnitude of the light vector. The curve of illumination distribution always has the property that points corresponding to the difference of oppositely directed radius vectors are found on a circle passing through P. The distribution curve of the illumination difference is a circle. The largest radius vector of this curve (the diameter of the circle) represents the projection of the light vector on the plane which intersects the brightness-distribution solid. In Fig. 11 the dotted line shows the illumination-difference distribution. To construct the illumination-distribution solid we must determine the light vector and one-half of the illumination-distribution solid; i.e., that half which is on one side of the plane of the illuminated element. The other half of the illumination-distribution solid can be computed, since the differences of opposite radius vectors are known.

Thus *an arbitrary illumination-distribution solid is not always possible.* The difference of any two opposite radius vectors should always represent the projection of the light vector; i.e., it should follow the cosine

law when we move from one direction to the next. The validity of the cosine law for the differences of any two opposite radius vectors is a *necessary* but not a *sufficient condition* that a curve represent the distribution of illumination.

To find the possible changes of illumination when the orientation of the element is varied, we shall need the illumination-distribution solid. The illumination of an element is determined by the brightness distribution for the rays passing through P. Let us consider a set of rays within a solid angle $d\omega$ and denote the normal illumination produced by them by dE_n. The resulting elementary illumination-distribution solid is a sphere with diameter dE_n. Constructing an illumination-distribution solid for each of the elementary solid angles, we obtain an infinite number of infinitesimal spheres of different diameters which have different orientations, which pass through the point P. The radius vector of the illumination-distribution solid is the sum of similarly directed radius vectors for these elementary spheres. *The limiting condition for the shape of the illumination-distribution solid is found in that each of the elementary solids is a sphere.* This must be the starting condition for the solution of the problem about the possible shape of the illumination-distribution solid. One must consider that each of the elementary illumination-distribution solids (the small spheres, each of which is produced by a single light ray) is found only on one side of the point (on one side from the plane, passing through the point, and normal to the considered ray). The radius vectors which are directed to the other side of this plane are equal to zero. *The cosine law may be applied only to a definite set of values of the angle between the ray and the normal to the illuminated element; for other values of this angle, the illumination is zero.* The resultant distribution of illumination is found by addition of curves which are similar to the one shown in Fig. 7. The curves have various amplitudes and are displaced with respect to each other in phase.

6. The Divergence of the Light Vector

The flux of the light vector,

$$F = \oint \mathbf{D} \cdot d\mathbf{s} \tag{63}$$

through a closed geometric surface is equal to the difference between the quantity of light entering the volume enclosed by the surface per unit time, and the quantity of light leaving the volume. The $d\mathbf{s}$

has the direction of the outer normal. Thus F represents the quantity of light produced (or absorbed) in unit time within the volume v. For a unit volume,

$$\frac{F}{v} = \frac{1}{v} \oint \mathbf{D} \cdot d\mathbf{s} \tag{64}$$

If the surface s is decreased so that the volume enclosed by it approaches zero, we obtain for the left side of Equation (64) the space density of light, produced per unit time, which we shall denote by f. If we deal with absorption rather than emission, the values of f will be negative. On the right side of Equation (64),

$$\lim_{v \to 0} \left[\frac{1}{v} \int_s \mathbf{D} \cdot d\mathbf{s} \right]$$

which is the *divergence of* **D**. Thus,

$$f = \text{div } \mathbf{D} \tag{65}$$

The space density of light, produced (or absorbed) per unit time, is equal to the divergence of the light vector.

From this follows that for all the points in the light field at which there is no emission or absorption of light,

$$\text{div } \mathbf{D} = 0. \tag{65a}$$

Thus in empty space the divergence of the light vector is equal to zero. The flux lines do not appear and do not disappear in that space, they may be grouped into flux tubes, and the field is solenoidal. Equation (65a) applies with extremely good approximation in the majority of problems in illuminating engineering, where the absorption of light in air may be neglected.

For empty space, therefore,

$$\text{div } \mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = 0 \tag{65b}$$

For a material medium, especially within light-emitting and light-absorbing bodies,

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = f \tag{65c}$$

where the function of position f is a measure of the generation or absorption of energy per unit volume surrounding the point in question. The quantity f may be either positive or negative.

Let us consider a finite volume v enclosed by the surface s. The flux of the light vector through the surface s is

$$\oint \mathbf{D} \cdot d\mathbf{s}$$

which, according to the law of conservation of energy, must be equal to the flux produced within the volume v, or

$$\int_v \operatorname{div} \mathbf{D}\, dv$$

The equation

$$\int_s \mathbf{D} \cdot d\mathbf{s} = \int_v \operatorname{div} \mathbf{D}\, dv \tag{66}$$

expresses Gauss's theorem: *The integral of a vector, taken over a closed surface, is equal to the integral of the divergence of this vector over the enclosed volume.*

7. Mechanical Analogies and the Geometric Interpretation

Let us consider some mechanical analogs of the light field produced by an arbitrary luminous body S (Fig. 20). The brightness B for rays that arrive at P from the body S may be different for different rays. Let us describe about the point P a sphere of unit radius. Rays from S cut out from the surface of this sphere the region ω. Let us ascribe to each element $d\omega$ of this area a weight number

$$dq = B\, d\omega$$

i.e., we shall consider the unit sphere as a material surface with the surface density equal at each place to the brightness B of the particular ray that intersects the surface of the sphere at that place. The space illumination is

$$E_0 = \int_\omega B\, d\omega = q$$

which may be considered as the weight of the shell of the sphere. The illumination of a surface s placed at the point P will be

$$E = \int_\omega B \cos\theta\, d\omega = \int h\, dq \tag{67}$$

where h is the distance along the normal from the element $d\omega$, which

has the weight dq, to the surface s. Thus the illumination is measured as a sum of the moments of the forces dq taken with respect to the illuminated plane. This sum is equal to the moment of the resultant force q which is applied to the center of gravity Q of the spherical shell.

When the point Q is found at a height h_0 with respect to the surface s, then

$$E = h_0 q.$$

Denoting the distance from the center of gravity Q to the illuminated point P by $r_0 (r_0 \leqq 1)$, we obtain

$$E = r_0 E_0 \cos \theta_0 \tag{68}$$

where θ_0 is the angle between the ray passing through the center of

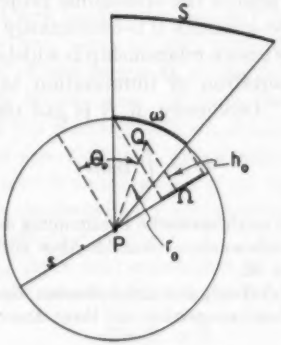


FIG. 20

gravity and the normal to the illuminated surface. Thus the light vector is equal to

$$\mathbf{D} = \mathbf{r}_0 E_0 \tag{69}$$

where $\mathbf{r}_0$ is the vector representing the distance from Q to P. The light vector **D** passes through the center of gravity, and its absolute value is

$$D = r_0 E_0 . \tag{70}$$

The quantity r_0, which is equal to the ratio of the numerical value of the light vector to the value of space illumination, characterizes the degree of diffusion. If the illumination is produced by a point source, $r_0 = 1$. Within a uniformly luminous body, $r_0 = 0$. This mechanical

analogy was first introduced by Mehmke.[21] It allows the solution of a number of photometric problems by applying known formulas of mechanics. For example, knowing that the center of gravity of a hemispherical shell is found at the middle of the radius ($r_0 = \frac{1}{2}$), we can say that the illumination within a uniformly luminous hemisphere of brightness B is equal to $\frac{1}{2} \cdot 2\pi B = \pi B$. The value of this mechanical analogy is found also in that it leads directly to the concept of the light vector.

The following geometrical interpretation of the values of illumination for uniformly bright surfaces is given by Wiener.[22] The illumination E is equal to

$$E = B\Omega \tag{71}$$

where Ω represents the area of the orthogonal projection of the spherical region ω (Fig. 20). The quantity Ω is customarily called the projection of the solid angle. The above relationship is widely used in photometry, and reduces the computation of illumination to a determination of areas of plane figures. Obviously, if B is not uniform,

$$E = \int_{\Omega} B \, d\Omega \tag{72}$$

[21] R. Mehmke, Über die mathematische Bestimmung der Helligkeit in Räumen mit Tagesbeleuchtung, insbesondere Gemäldesälen mit Deckenlicht. Zs. für Math. u. Phys., **43**, 1898, p. 41.

[22] Christian Wiener, Lehrbuch der darstellenden Geometrie, Leipzig, 1884. (Vol. 1, Chapter on "Beleuchtungslehre mit ihrer Anwendung auf ebenflächige Körper.")

CHAPTER V

The Structure of the Light Field

1. Lines of flux

Generally to each point of the light field corresponds a definite value and direction of the light vector. Thus one may construct lines of the light vector or *flux lines*. A flux line is a space curve, the tangents to which coincide with the directions of the light vector.

Consider a point P to which belongs the light vector **D**, and assume that P is located on the surface of a luminous body. Then draw through P a straight line having the length PP_1 in the direction of the vector **D**. To this line segment PP_1 add another straight section P_1P_2 in the direction of the vector **D** at the point P_1. Repeating this process, we obtain a broken line. If now the size of the line segments is reduced, in the limiting case we obtain a curve, each element of which coincides in direction with the direction of the light vector at that point. This curve is called a *line of the light vector* or a *flux line*. The light vector is a single-valued point-function: Through each point of the field passes a single flux line.

For a region of the light field where div $\mathbf{D} = 0$, the flux line passes uninterrupted. At the source of the light field,

$$\text{div } \mathbf{D} > 0$$

while at a sink of the light vector, where the light is absorbed,

$$\text{div } \mathbf{D} < 0.$$

In empty space, the flux lines emerge from the surfaces of luminous bodies and pass to infinity or are interrupted by absorbing bodies.

Strictly speaking, a flux line is not interrupted exactly at the surface of the illuminated body, but at some depth in it. The line may be followed the deeper, the more transparent is the illuminated body. Similarly, a line is not generated on the surface of a luminous body, but is generated within its depths, where the process of light emission takes place. If the coefficient of absorption for its own radiation is very large, as for example in the case of incandescent metals, all the radiation that emerges is produced by surface layers. In the general

case, however, an element of the outer surface of a luminous body should be considered as a window through which the inner radiation emerges. This becomes obvious for a light field produced by a gaseous-discharge lamp. The flux line originates not at the surface of the envelope, but in each element of the volume of the gas. The sources of the light vector are identical with the light-emitting centers. The sinks of the vector coincide with absorbing centers. Thus in a light-absorbing medium, flux lines emerging from a luminous body will be thinned out gradually as they proceed away from the source; for the

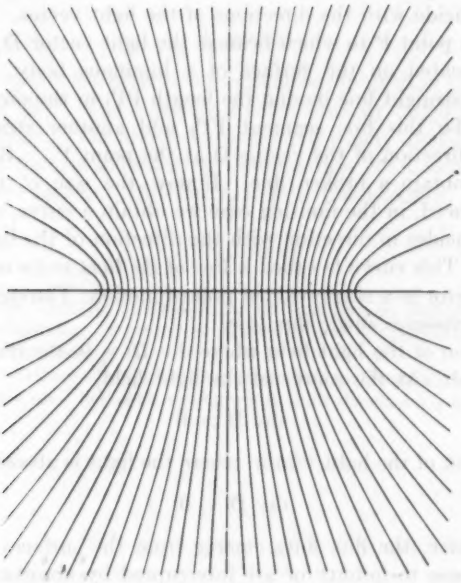


FIG. 21

light lines must terminate again and again as light is absorbed. If the medium emits as well as absorbs light, the sources and sinks of the light vector are distributed over the whole body.

A picture showing the flux lines illustrates the structure of the light field. Let us consider a number of examples. The light field produced by a point source is radial, corresponding to the field of a positive charge of electricity or of a positive magnetic pole. Fig. 21 shows the flux

lines produced by an infinite, uniformly luminous strip.[23] Here the lines are confocal hyperbolas. Inasmuch as the concept of the line of luminous flux is a novel one, we may mention once more that a difference exists between a *flux line* and a *light ray*. Through each point of the field pass a multitude of light rays, each of which is a straight line. The concept of a ray does not require any explanation, but the new concept of the flux line is not so easily visualized.

In concluding this section, we note that the shape of the flux line may be identical for different light fields. This is always the case when the fields differ from each other only in the numerical values of the light vectors. Also, for example, for a uniformly luminous, infinite cylinder, as well as for a point source, the flux lines are radially diverging rays. The flux lines emerging from a circular disc, as well as those from a luminous, infinite strip (Fig. 21) are represented by confocal hyperbolas. Let us also recall that the light vector defines directly not the condition of illumination, but the transfer of energy. From the standpoint of energy transfer, the fields may be identical while at the same time *the brightness distribution solids and the corresponding illuminations may be quite different for the two cases.*

2. Tubes of Flux

We may construct an infinite number of flux lines in the light field; but it is natural to introduce the concept of *tubes of flux*, since the replacement of the infinity of lines by a finite number of tubes results in a convenient method of approximate study of the light field.

Let us consider a section of a light tube, bounded by two arbitrary surfaces s_1 and s_2 (Fig. 22), and write for this section the equation of energy balance. The sides of the tube are illuminated from the inside and from the outside and these illuminations are equal for each element of the side surface. Indeed, the sides are formed by flux lines, so the projection of the light vector on the normal to the element considered, which projection of the vector determines the difference of illumination on the two sides of the surface element, is zero for all the points of the side surface of the tube. Thus the light flowing in per unit time through the side surface of the tube is equal to the light flowing out through that surface.

Denoting by F_1 the flux through section s_1 and by F_2 the flux through the surface s_2, we have, according to the law of conservation of energy,

[23] Edward P. Hyde, Geometrical Theory of Radiating Surfaces with Discussion of Light Tubes. Bu. Stds., Bull. 3, 1907, p. 81.

$F_2 = F_1 + F_v$. Here F_v is a measure of the intensity of the source (or the sink) within the volume v of this section of the tube. In case we consider a section of the light tube in empty space, the volume v does not contain either sinks or sources, and $F_v = 0$ and $F_2 = F_1$. Because luminous flux is the flux of a vector, the above equality follows immediately from a theorem of the vector field: *the flux of a vector through a vector tube containing no sources or sinks is constant.* This statement represents a special case of Gauss's theorem. Thus in empty space the luminous flux F through an arbitrary cross-section of the light tube is always the same, F = const. For any arbitrary cross-section of the tube, the difference between the fluxes, on this section from both of its sides, remains constant. *This constant characterizes the particular tube.*

If the light field is produced by a convex luminous body, the tube constant has a simple physical meaning. Obviously, the tube constant is equal to the flux F emitted in all directions by the surface

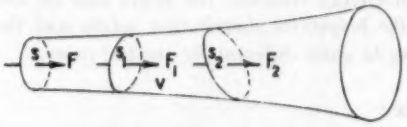


FIG. 22

element S of the luminous body, (Fig. 22). Thus in photometric computation it may be assumed that all energy emitted by a portion of a convex luminous surface flows along a light tube. This is, of course, a mathematical abstraction, as in reality the area S of the emitting surface sends light in all directions within a solid angle 2π. The light falls on section s of the tube, generally, from all possible directions and the surface s is illuminated as a rule from both sides. In a number of cases, however, it is convenient to use this abstraction and to assume that all the light emitted by the area S of the luminous surface flows along the tube, entering the volume v (Fig. 22) through the section s_1 and flowing out of it through the surface s_2.

On the basis of these formal concepts, it is easy to compute in a number of cases the luminous flux from one surface to another. For this purpose, we draw flux lines through the points of the boundary of the illuminated surface. These lines cut out from the luminous surface a definite area. The flux emitted by this area in all directions will be equal to the flux from the illuminated surface, or will be equal to the

difference of the fluxes on the two sides of it if the surface is illuminated from both sides. This leads to a simple method of graphical or analytical computation of the flux from one surface to another. We shall use this method further when treating the light field produced by a luminous disc and by a luminous strip.

To illustrate the vector method, consider one of the photometric problems of Lambert: "Required the flux from a uniformly luminous sphere to a circular disc, the axis of the disc passing through the center of the sphere." (Fig. 23.) From considerations of symmetry, it is obvious that the flux lines are straight and diverge radially from the center of the sphere. Let us draw flux lines through the circular boundary of the illuminated disc. The light tube formed by these

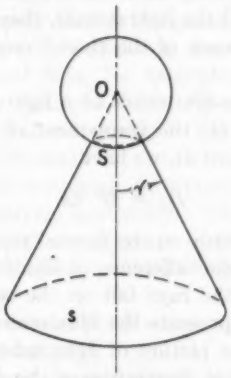


FIG. 23

lines has the shape of a circular cone and cuts an area S from the surface of the luminous sphere. According to what has been said, the flux F on the disc is numerically equal to the flux from the spherical surface S. This flux is obviously proportional to the total flux F_0 emitted by the sphere, the constant of proportionality being given by the ratio of the area S to the total area of the sphere. Thus,

$$F/F_0 = \omega/4\pi$$

or

$$F = F_0 \frac{\omega}{4\pi} = \frac{1 - \cos\gamma}{2} F_0 \qquad (73)$$

The sphere may be replaced by a point source located at its center, which is permissible in all cases when the plane of the illuminated surface does not intersect the sphere. More difficult examples which we shall encounter later will convincingly illustrate the practical usefulness of the vector concepts in photometry.

Let us consider now in greater detail some properties of the tubes of the light vector. The part of space in which we study the light field may be subdivided into a number of light tubes. To each of these tubes will correspond a definite value of luminous flux. It is natural to subdivide the region in such a manner that for all the tubes this flux will be the same. It is obvious that the space may be subdivided into a given number of tubes in a large number of different ways. When the space is subdivided into an infinite number of elementary light tubes, which we shall call the *light threads*, they are sufficiently narrow so that the field within each of the thread cross-sections may be considered uniform.

Let us consider a cross-section ds of a light thread that carries the flux dF. Denoting by D_n the component of the light vector along the normal n of the element ds, we have

$$D_n = dF/ds. \tag{74}$$

The projection of the vector on the normal represents the flux density through the surface or the difference of illumination of the two sides of that surface. If all the rays fall on the section ds from one side of it, the flux density represents the illumination of the element ds.

We have seen that the picture of light tubes intersecting a surface not only yields a graphical illustration of the distribution of luminous flux, but allows us to determine numerically the *flux density* of different parts of that surface, as well as the *flux* through an arbitrary part of the surface. This computation will be the more accurate the larger is the number of light tubes considered. In the limiting case, when we change from light tubes to light threads, we can determine the flux density as a function of position on a surface, and can determine the exact value of the flux through any contour on this surface.

Let us replace each of the light tubes by a single line within the tube, and assume that the entire flux of the tube is along this line. A surface placed in the light field will be pierced at a number of points by the *axial flux lines*. The number of such points in a portion of the surface is proportional to the luminous flux through this area. The number of points per unit area determines the flux density. If the

surface is illuminated from one side only, the distribution of illumination on the surface is given by the flux diagram and the determination of the flux is reduced to counting the number of points within the given contour.

In the following paragraphs we shall treat several examples of the vector method of computation and the graphical representation of the light field. These examples will illustrate the soundness of introducing Faraday's concept of a force-field into photometry. Before we consider these examples let us say a few words about the possibilities of graphical representation of the vector field.

As a rule, the flux lines are represented by space curves that cannot be shown graphically because they are not in the plane of the drawing. In a number of cases, however, the flux lines are represented by plane curves which may be subdivided into families, each of which is contained in a single plane. This situation occurs in

a) A two-dimensional field; for example, the field from an infinitely long, uniformly luminous strip. The flux lines are contained in planes that are normal to this strip. The intersections of the field by these planes are identical.

b) A field with axial symmetry; for example, a light field produced by a circular disc of uniform luminosity. The flux lines are contained in meridional planes and the diagram of flux lines is the same for all these planes. In the same class belongs a field produced by a luminous straight line.

c) A field resulting from a point source, where the flux lines coincide with the radius vectors.

We have indicated three possible classes of light fields which permit simple graphical representation. In some special cases, the light field may simultaneously belong to two classes. For example, the light field produced by a uniformly luminous circular cylinder of infinite length is a plane field, and on the other hand it is a field possessing axial symmetry. The light field at a sufficiently large distance from a symmetrical source possesses axial symmetry and may also be regarded as a field of a point source.

3. The Light Field of a Strip

Let us begin with the two-dimensional case of an infinitely long, perfectly diffusing, uniformly luminous strip (Fig. 24). The strip extends to infinity on both sides of the plane of the drawing. We find in the plane of the drawing a family of flux lines whose equation is

easily found. Let us consider a point P_1 of the light field. The brightness on all the rays arriving at the point P_1 from the strip is equal. These rays form a symmetrical pencil, and the line of symmetry bisects the angle FP_1F', which we shall denote by α. It follows that the light vector **D** is directed along the bisectrix of α. Thus the line of the vector **D** (the flux line) must possess the property that its tangent at any point P_1 bisects the angle FP_1F'. This condition, as it is well known from analytic geometry, is satisfied by an hyperbola, the foci of which are at F and F′. Thus the flux lines in the plane of the drawing form a family of confocal hyperbolas. All the planes that are parallel to the plane of the drawing, at a finite distance from it, yield a similar picture. Thus the field may be subdivided into light tubes by a family of planes and by a family of confocal hyperbolic cylinders. Each pair of adjacent cylinders, intersecting a pair of ad-

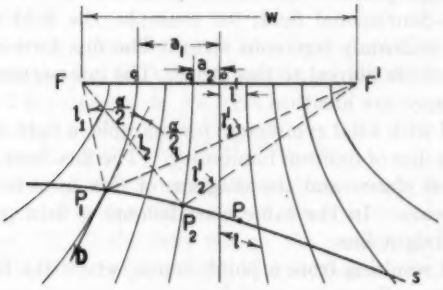


FIG. 24

jacent parallel planes produces a light tube. In order that the luminous flux through a cross section of a tube be the same for all the tubes it is necessary and sufficient that the tubes originate from equal areas of the luminous surface. For this purpose, we subdivide FF′ into equal sections and pass the hyperbolas through these points. The diagrams of Figs. 24 and 21 allow us to judge not only the direction of the vector but also the flux density D_n at any point P of a surface S. To determine D_n we draw the intersection of this surface with the plane of the drawing and measure along the line of intersection the distance l between the two adjacent flux lines forming the tube containing the point P (Fig. 24). Denoting by l' the smallest distance between two neighboring lines at the place where they emerge from the luminous strip, and by L the luminosity of the strip,

$$D_n/L = l'/l.$$

If light falls on only one side of the illuminated surface, then D_n represents the illumination of that surface, while D_n/L represents the coefficient of illumination. This coefficient gives the ratio of the flux density on the illuminated surface to the flux density at the source. From Fig. 24 we can judge the magnitude of the coefficient of illumination at various points of any surface in the field.

The number of light tubes arriving at a given area of the surface gives the luminous flux on this area. Let us derive, on the basis of vector concepts, a formula for the value of the luminous flux F on the region P_1P_2 of the surfaces (Fig. 24). The total flux emitted by the strip from one side is denoted by F_0. The flux lines through the points P_1 and P_2 emerge from the points O_1 and O_2 of the luminous strip and have the distances a_1 and a_2 from the center line of the strip.

Thus,

$$F/F_0 = \frac{a_1 - a_2}{w},$$

where w is the width of the luminous strip.

This formula reduces the computation of luminous flux to a construction of two hyperbolas and a determination of the shortest distance between them. Even this construction may be avoided, for

$$2a_1 = l_1' - l_1; \qquad 2a_2 = l_2' - l_2.$$

Thus,

$$F/F_0 = \frac{1}{2w}[(l_1' - l_1) - (l_2' - l_2)] \tag{75}$$

This formula yields the simplest method of computing the luminous flux from one infinite strip to another.

The fact that the flux lines are hyperbolas allows the easy determination of the magnitude of the light vector. To this end, we consider two infinitely close confocal hyperbolas, the shortest distance between which is da (Fig. 25). It is easily shown that the length $d\lambda$ of the element of the orthogonal trajectory enclosed between the two hyperbolas is

$$d\lambda = \frac{d\alpha}{\sin \alpha/2} \tag{76}$$

It follows that the magnitude of the light vector **D** is

$$D = L\frac{da}{d\lambda} = L \sin \frac{\alpha}{2} \tag{77}$$

Thus the light vector bisects the angle subtended by the luminous strip, and the magnitude of the vector is proportional to sin $\alpha/2$. This may be shown also by starting from the geometrical interpretation given in Section 7 of Chapter IV. At large distances from the luminous strip, the field is a radial one with the asymptotes of the hyperbolas as flux lines. In these regions of the light field, the illumination is inversely proportional to the first power of the distance. The magnitude of the light vector is

$$D = \frac{Bw \cos \theta}{l} \tag{78}$$

where B is the brightness of the strip, l the distance from the strip,

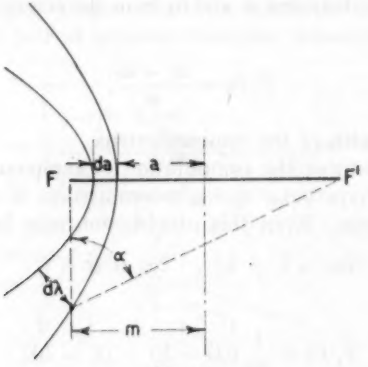


FIG. 25

and $w \cos \theta$ is the projection of FF′ on a straight line normal to the direction considered, which in turn forms the angle θ with the axis.

The case of a uniformly luminous strip is of practical value in the computation of natural lighting.[24] As an example, let us consider the light field, produced by the diffused light of the sky, between the walls of equally tall houses bounding a street. We assume that the street is infinitely long, and that the sky is perfectly diffusing. Then the field between the walls of the houses will be the same as that produced by an infinitely long source located above the street and having the same luminosity as the sky. The flux lines give not only a qualitative but also a quantitative picture of the distribution of daylight on the walls

[24] A. Gershun, Computation of Natural Lighting. Trans. Opt. Inst., 5, No. 44, 1929. Leningrad.

of the houses and on the surface of the ground.* A more detailed treatment of vector methods of computing natural lighting is beyond the scope of this book but is considered in a paper by the author.

4. The Light Field of a Luminous Disc

As an example of a field possessing axial symmetry, let us consider the illumination produced by a uniformly luminous, perfectly diffusing circular disc. The solid angle subtended by the disc at a point P has the shape of an elliptic cone. In Fig. 26, FF′ is the diameter of the disc and FPF′ is the axial cross-section of the cone. The axis of the cone, along which the light vector is directed, is the bisectrix ZP of the angle α; the point M in which ZP intersects the disc represents its "light-center" for the point P.

Thus the flux lines are again represented by confocal hyperbolas with the foci at FF′. Rotating these hyperbolas about the axis of the disc we obtain hyperboloids of revolution that subdivide the whole space into light tubes. In order that these tubes shall carry equal fluxes, the annular regions of the disc from which they originate must be equal in area. For this we subdivide the diameter of the disc FF′ into sections such that the squares of the distances from the center form an arithmetic progression.

In this example, the flux diagram does not characterize the variation of the light vector as simply as in the two-dimensional example. Indeed, for the two-dimensional field the magnitude of the light vector was fully determined by the length $d\lambda$ of a section of the orthogonal trajectory enclosed between two neighboring flux lines (Fig. 25). For the disc, the value of the light vector varies inversely as the area of the conical surface formed by $d\lambda$ when rotated about the axis of the disc; or D is inversely proportional to $md\lambda$, where m is the distance from the point to the axis of the disc. The distance between neighboring lines characterizes directly the distribution of illumination only when m is constant, i.e. for all cylindrical surfaces having a common axis with the disc.

To determine the magnitude and direction of the light vector, let us consider a sphere determined by the boundary of the luminous disc and the point P (Fig. 26). A sphere possesses the property that for all its points the angle α is the same. Obviously, the inner surface of the sphere is evenly illuminated, since the light field of the disc S is identical

* [But not inside the rooms.] Trans.

with the light field produced by the cap σ. The area of the spherical cap is

$$\sigma = 2\pi R^2 (1 - \cos \alpha) \tag{79}$$

where R is the radius of the sphere. Attributing to the cap a luminosity L, which is equal to the luminosity of the disc, we find that the flux emitted by σ is $L\sigma$. This flux is uniformly distributed over the whole surface of the sphere* and produces an equal illumination E at all points, or

$$E = \frac{L\sigma}{4\pi R^2} = L \frac{1 - \cos \alpha}{2} = L \sin^2 \frac{\alpha}{2} \tag{80}$$

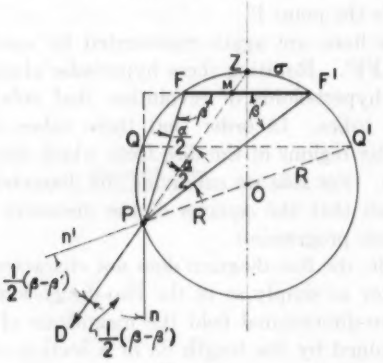


FIG. 26

It is of interest to note that the sphere described about the disc represents not only a uniformly illuminated surface but also the geometrical locus of all points in which the illumination of an element parallel to the plane of the disc is constant and equal to E. Indeed, the light vector at P is directed along the bisectrix of the angle α. This bisectrix forms equal angles with the normal to the illuminated element when it is an element of the sphere (normal n'), as well as when it is parallel to the plane of the disc (normal n). Thus the illumination E on a plane that is parallel to the disc is

$$E = E' = L \sin^2 \frac{\alpha}{2}. \tag{81}$$

* [Note that in the entire discussion the assumption is tacitly made that the sphere reflects no light. In an *actual* sphere, the illumination may be much greater than that given by Eq. (80) because of interreflections.] *Trans.*

Let us now introduce the angles β and β' between the outer rays FP and F'P and the plane of the disc. The angle between **D** and the normal n is

$$\frac{\beta - \beta'}{2}$$

Thus the absolute value of the light vector is

$$D = \frac{L \sin^2 \alpha/2}{\cos \dfrac{\beta - \beta'}{2}} \tag{82}$$

This formula, together with the condition that the light vector is directed along the bisectrix of the angle FPF', completely defines the illumination produced by the disc. The illumination is given by the projection of **D** on the normal to the illuminated surface. Let us introduce an angle defined by the relationship

$$\sin \frac{\alpha'}{2} = \frac{\sin \alpha/2}{\cos \dfrac{\beta - \beta'}{2}} \tag{83}$$

Then the absolute value of the light vector is

$$D = L \sin \frac{\alpha}{2} \sin \frac{\alpha'}{2} \tag{84}$$

The angle α' is the angle subtended by the chord which passes through the light center M and which is normal to the diameter FF'. The conical surface enclosing the rays passing from the disc to the point P is an elliptical one, and it cuts a spherical ellipse from a unit sphere with center at P. The minor and major semi-axes, expressed in angular units, are $\alpha/2$ and $\alpha'/2$. The orthogonal projection of this spherical ellipse on a surface that is normal to the axis of the cone represents a plane ellipse with semi-axes of sin $\alpha/2$ and sin $\alpha'/2$. The area of the ellipse is

$$\Omega = \pi \sin \frac{\alpha}{2} \sin \frac{\alpha'}{2}. \tag{85}$$

Multiplying the expression for Ω by the brightness B, we obtain, according to Section 7 of Chapter IV, the absolute value of the light vector. For the points on the axis of the disc, $\alpha = \alpha'$, and

$$D = L \sin^2 \frac{\alpha}{2} \tag{86}$$

The illumination varies inversely as the square of the distance from the rim of the disc. This problem has been treated in detail by Steinmetz[25] and by Yamauti.[26]

At points that are close to the disc, the illumination does not depend on the distance from the disc. In this region, the flux lines are normal to the plane of the disc. The cross-section of the light tubes (and therefore, also the illumination) does not vary as we move from one point to another and the light field is homogeneous. At great distance from the disc, the light field is a radial one.

The circular disc has always been a favorite example in photometric computations and is considered in the books of Lambert and of Beer, as well as in a number of other publications. The formulas expressing the illumination are usually derived as a result of tedious integrations over the surface and are usually presented in a clumsy and complicated way. It is often forgotten that the law of similitude may be exactly applied in photometry. *The scale is arbitrary because the linear dimensions appear in all the formulas in the form of their quotients; and the photometric relationships become elegant, and thus convenient to remember and to use in computations, only when the whole geometry is characterized by angles.*

The problem of the illumination produced by a uniformly luminous circular disc also gives the inverse problem of finding the flux from an element of perfectly diffusing surface to a circular disc which is placed arbitrarily with respect to that element. This follows from the principle of reciprocity which states that if two bodies are equally bright, then the flux from the first body to the second is equal to the flux from the second body to the first. The correctness of this statement follows directly from the formula for the luminous flux from an element of one surface to an element of another surface (Section 1, Chapter 2). This statement also follows from the second law of thermodynamics.

Lambert, having proved the principle of reciprocity (Theorem 16 of "Photometry"), considers its use in photometric computations. For example, using the formula for the illumination of a plane from a disc which is parallel to it, we find according to the principle of reciprocity

[25] Chas. Steinmetz, Radiation, Light and Illumination. McGraw-Hill Book Co., New York, 1918.

[26] Ziro Yamauti, Geometrical Calculation of Illumination due to Light from Luminous Sources of Simple Forms. Researches of the Electrotechnical Laboratory, Ministry of Communications, Tokyo, No. 148, 1924.

that the perfectly diffusing element dS of brightness B sends the flux dF:

$$dF = B\pi \sin^2 \frac{\alpha}{2} dS \tag{87}$$

to a disc parallel to dS. Here α is the angle subtended by the disc from the element dS. This formula, which usually is proved only for the special case of the luminous element on the axis of the illuminated disc, is of considerable importance in optical apparatus. It yields an expression for the luminous flux entering an optical apparatus from an element normal to the axis of the apparatus.

Let us now compute Lambert's problem of the flux from a luminous circular disc FF′ to the coaxial disc PP′ of Fig. 27. (Theorems 17 and 19 of "Photometry"). This flux must be equal to the flux emitted in all directions by the inner part of the disc of radius a, where a is the

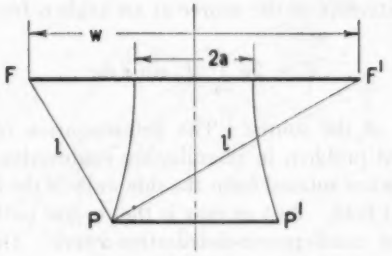


FIG. 27

real semi-axis of hyperbolas that pass through the rim of the illuminated disc. Denoting by F_0 the total luminous flux from the disc FF′, and by F the part on the illuminated disc PP′, we have

$$F/F_0 = \left(\frac{2a}{w}\right)^2 = \left(\frac{l' - l}{w}\right)^2 \tag{88}$$

In conclusion, we note that the light vector within a uniformly luminous hollow space with a circular opening may be reduced to the case of the circular disc. If the space were completely closed, we would have the case of "neutralization", and the light vector would be a null vector at all points within the enclosure. When an opening is present, the light vector is equal in magnitude to the vector which the opening would produce, were it a source with the same brightness as the walls of the enclosure. The direction of this vector is opposite to the one which would be produced by the opening.

5. The Light Field Produced by Point Sources

The flux lines of a point source are radially diverging straight lines. Subdividing the space into a set of cones, each with its apex at the source and each carrying equal luminous flux, we obtain tubes of equal flux. This construction can obviously be carried out in a large number of ways.

If the source has a symmetrical distribution about an axis, the space may be subdivided into tubes by conical surfaces, the axes of which correspond to the axis of symmetry. When n light tubes are constructed, the angle α_i between the i^{th} conical surface and the axis is determined by the relationship:

$$2\pi \int_0^{\alpha_i} I_\alpha \sin \alpha \, d\alpha = \frac{i}{n} F \tag{89}$$

where I_α is the intensity of the source at an angle α from the axis, and

$$F = 2\pi \int_0^{\pi} I_\alpha \sin \alpha \, d\alpha \tag{90}$$

is the total flux of the source. The determination of the angles α_i represents a usual problem in illuminating engineering computations. The rays, which when rotated form the side walls of the light tubes, give a map of the light field. Just as easy is the reverse path from the map of the field to the candlepower-distribution curve. Dividing the flux through a single tube by the solid angle of the tube, we obtain the average intensity in the directions contained within that tube.

The construction of the map of the light field simplifies illuminating-engineering computations in a large number of cases. An illustration is the vector method for computing the candlepower distribution from a symmetrical specular reflector, as suggested by the author and elaborated in collaboration with N. G. Boldyreff.[27] Almost as easy is the reverse problem of a design of a reflector to produce a required transformation of the light field of the lamp.

Let us now treat a somewhat different method of graphical representation of the light field from a point source. This method was discovered by the author in collaboration with N. G. Boldyreff and may be also used for sources with asymmetric distributions.

The construction of the flux lines in a plane S containing the source I (Fig. 28) is based on the following principle. Let us consider the

[27] A. A. Gershun and N. G. Boldyreff, The Vector Method of Computation of Symmetrical Reflectors Used in Illumination. "Svetotechnika", 1936, No. 1.

plane S′, which is infinitely close to S and which intersects the latter along the straight line C_o. To characterize the distribution of the flux emitted by the source within the angle formed by the planes S and S′, we subdivide this angle into tubes carrying equal fluxes by planes Q_i which are normal to S. The following relationship must be fulfilled:

$$2\pi \int_{\beta_i}^{\beta_{i+1}} I_\beta \sin \beta \, d\beta = \text{const.} \tag{91}$$

where I_β is the intensity of the source in the direction contained in the plane S and making the angle β with the straight line C_o. The diagram of rays obtained by the intersection of the plane S by the planes Q_i possesses the following important properties. Let us consider a plane s which intersects the plane S along a straight line C which is parallel

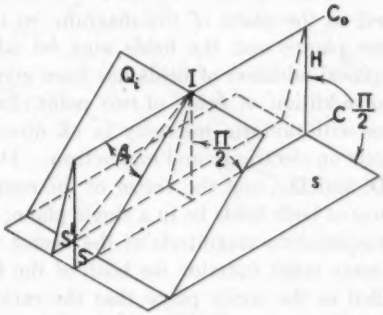


FIG. 28

to C_o and passes at a distance H from the latter. The light rays of the vector map will cut the straight line C into a number of sections, whose length will directly characterize the distribution of illumination in the plane s along the straight line C. Indeed, the illumination of the plane s at points on C is inversely proportional to H and inversely proportional to the distances intercepted on C by neighboring rays of the vector diagram.

6. Addition of Fields

When there are a number of luminous bodies in the field, each of the quantities characterizing the light field represents a sum of values corresponding to each of the luminous bodies separately. *The light field produced by a set of luminous bodies is the sum of the light fields produced by these bodies individually.*

Thus we arrive at the question of the addition of fields. This question is of great practical importance, inasmuch as in practice we always encounter lighting from a set of sources, and all objects in the field reflect to some extent. In some cases, the source is complex and must be divided into a number of simpler elements. When adding the fields of several sources, one must take into account the production of shadows caused by the screening action of one luminous body with respect to a light field produced by another source. Sometimes the effect of the shadow may be accounted for if we introduce not only the addition of fields but also their subtraction.

In a number of cases (a two-dimensional field, a field with axial symmetry, a field of a set of point sources on a straight line) the fields that are added, as well as the resultant field, may be represented graphically. Cross-sections are considered such that the vector has zero component normal to the plane of the diagram, so the flux lines are contained in these planes and the fields may be added graphically. A method of graphical addition of fields has been given by Maxwell.[28]

The problem of addition of fields of two point charges (in photometry, point sources with uniform intensity in all directions) is treated in a number of texts on electricity and magnetism. Denote the vectors to be added by $\mathbf{D}_1$ and $\mathbf{D}_2$, and the vector of the resultant field by $\mathbf{D}$. Let the vector lines of both fields be in a single plane; and, in addition, let the ratio of the absolute magnitude of the vector to the density of the lines be the same point function for both of the fields. Then the condition is fulfilled in the entire plane that the ratio of the absolute value of the vector $\mathbf{D}_1$ to the absolute value of the vector $\mathbf{D}_2$ is equal to the ratio of the density of vector lines of the first field to the density of vector lines of the second field.

The vector lines of both fields divide the plane into a number of regions (Fig. 29) and define two new sets of curves, which are directed along the diagonals. Maxwell showed that the fields of the diagonal curves completely determine the field of the resultant vector. Depending on the choice of one of the two systems of diagonals, we obtain the field of the vector $(\mathbf{D}_1 + \mathbf{D}_2)$ or of the vector $(\mathbf{D}_1 - \mathbf{D}_2)$. The correctness of Maxwell's theorem, upon which the graphical addition of vector fields is based, may be easily shown. Let us consider a very small region in the plane, such that within that region the lines of each

[28] James Clerk Maxwell. A Treatise on Electricity and Magnetism (Vol. I, Part I, Electrostatics. Ch. VII.) See also the footnote by A. Cornu to this chapter in the French edition.

of the two vectors $\mathbf{D}_1$ and $\mathbf{D}_2$ may be considered as equidistant parallel lines. These subdivide the regions of the plane into a number of parallelograms. We denote by l_1 and l_2 the distances between neighboring lines of the component fields, and by l the distance between the diagonal lines of the resultant field. Consider the parallelogram, one corner of which is at P, and denote by $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}$ the vectors corresponding to the sides of the parallelogram and its diagonal. A comparison of the three expressions for the area of the parallelogram gives

$$e_1 l_1 = e_2 l_2 = el \tag{92}$$

$$e_1 : e_2 : e_3 = \frac{1}{l_1} : \frac{1}{l_2} : \frac{1}{l} \tag{93}$$

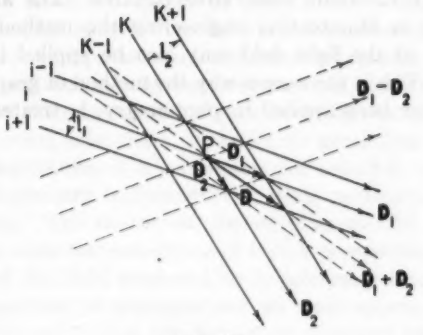


FIG. 29

According to our assumption,

$$D_1 : D_2 = \frac{1}{l_1} : \frac{1}{l_2} \tag{94}$$

Therefore, $\mathbf{e}_1$ and $\mathbf{e}_2$ represent on some scale the vectors $\mathbf{D}_1$ and $\mathbf{D}_2$. The vector $\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2$ represents on the same scale the vector $\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2$. The parallelogram formed by sections of vector lines meeting at the point P, represents the parallelogram of addition of vectors.

Vector fields produced by several sources of light may have a common axis of symmetry, may be two-dimensional fields, or may be fields of point-sources along a straight line. On one drawing we superpose the maps of two of the fields to be added, making use of the requirement that the number of flux lines of each field should correspond to the

strength of its source. Having numbered the lines, we connect by curves the points of intersections of the lines of both fields in the following manner. Let i be the order number of a line of the first field and k a corresponding number in the second field (Fig. 29). Each point of intersection is characterized by two indices and may be denoted by the symbol (i, k). The transfer from the point (i, k) to the point $(i + 1, k + 1)$ is in the direction of the resultant vector. Connecting by a curve the points: $\cdots (i - 1, k - 1); (i, k); (i + 1, k + 1); \cdots$ we obtain a line of the resultant field. Adding a field obtained in this manner to the field of a third source, we obtain the field of the sum of three vectors, and so on.

In physics, the graphical representation and addition of fields is used commonly only to obtain illustrative pictures. The author believes, however, that in illuminating engineering the methods of graphical representation of the light field may also be applied in quantitative design work. This is the reason why the method of graphical addition, which has never been applied in photometry, is treated here in such detail.

CHAPTER VI

Photometric Computations

1. General Theorems

In the present chapter we shall consider photometric computations for large luminous surfaces. The assumption will be made that the surfaces have uniform luminosity and that they satisfy Lambert's law. The light field produced by such surfaces is subject to a number of simple relationships which serve as a basis for photometric computations. We shall treat the theory in general terms and shall avoid tedious computations which often have the character of exercises in integral calculus. Computation of the illumination and the flux from large luminous surfaces usually requires multiple integrations in each special case, but these integrations may be performed once for all in deriving the general formulas. In order to generalize and simplify photometric computations it is permissible and sound to create a special mathematical machinery by introducing, when necessary, new mathematical concepts. The history of the development of physics illustrates on a large scale the correctness of such a procedure.

The theory of the field produced by a uniformly luminous surface is primarily a problem of geometry and we shall approach it as such. Let us consider a point P of the light field produced by a uniformly luminous surface S bounded by the contour C, the brightness of this surface being denoted by B. The surface S subtends the solid angle ω at the point P. The angle ω is composed of elements $d\omega$ subtended by the elements dS of the luminous surface. To characterize the light field, we shall also introduce a vector solid angle $d\boldsymbol{\omega}$. The space illumination is

$$E_0 = \int_\omega B \, d\omega \tag{95}$$

and the light vector is

$$\mathbf{D} = \int_\omega B \, d\boldsymbol{\omega} \tag{96}$$

where the symbol ω below the integral sign indicates that the integration is performed within the solid angle ω.

For a uniformly luminous surface, B is the same for all the rays at the point P within the solid angle ω. Taking B outside the integral sign, we obtain

$$E_0 = B \int d\omega \tag{95a}$$

and

$$\mathbf{D} = B \int_{\omega} d\omega \tag{96a}$$

The integral in Eq. (95a) represents the solid angle ω. Thus the computation of space illumination in this case is reduced to the geometrical problem of computing the solid angle. The expression obtained for the space illumination,

$$E_0 = B\omega \tag{97}$$

is analogous to the expression for the potential produced by a magnetic sheet.

Analogously, the computation of the light vector may be reduced to a purely geometric problem. For this we shall introduce a new mathematical concept of the vector ω' obtained by the summation of vectors $d\omega$:

$$\omega' = \int_{\omega} d\omega \tag{98}$$

the integration being performed over all directions of space within the angle ω.[+] To each conical surface corresponds a definite solid angle ω

[+] [Note that $|\omega'|$ does not give the value of the solid angle ω, since the elementary vectors $d\omega$ are not collinear.] Trans.

as well as a definite vector ω'. The reverse statement is obviously not true, for to one value of ω or ω' may correspond different conical surfaces.

The light vector is defined as

$$\mathbf{D} = B\,\omega' \tag{99}$$

The absolute value of the vector ω' is denoted by ω'. It obviously depends not only on the magnitude ω but also on the shape of the conical surface enclosing the solid angle; and $\omega' \leqq \omega$, the equality sign being for the case of an infinitely small solid angle. When the contour C of the luminous surfaces is given, then to each point of the field cor-

responds a definite cone. This cone has its apex at P and has as its base the contour C. To each conical surface, and thus to each point of the field, correspond definite and unique values of ω and ω'. The study of the light field is completely reduced to the study of two geometrical fields: the field of the *scalar solid angle* and the field of the *vector solid angle*. The space density of radiant energy is determined by the scalar solid angle at a given point of the field. The flux lines are the lines of the solid-angle vector, and a computation of the flux through a surface is reduced to a computation of a flux of the solid-angle vector.

The following interpretation of ω' may be given. Let us describe a sphere of a unit radius about the apex of the conical surface considered. The conical surface will cut from this sphere a region whose area is equal to ω. To each elementary solid angle $d\omega$ will correspond a definite element of the spherical surface. The vector corresponding to this element is $d\omega'$, which is directed along the inner surface normal. Thus the vector ω' may be considered as the vector sum of the elements of the surface cut from a unit sphere by a given cone, or, using the terminology of vector analysis, as the vector of this surface. It is easy to show that the vector ω' passes through the center of gravity of the surface.

The projection of the solid-angle vector on a given direction is equal to the area of the projection, on a plane which is normal to the given direction, of the region cut from the unit sphere by the cone. Thus if the direction of the vector is known (for example, from considerations of symmetry) then its magnitude may easily be determined as the area of the projection of a spherical region on a plane that is normal to the direction of the vector.

The solid-angle vector represents a fundamental quantity in the computation of illumination produced by large luminous surfaces of uniform brightness. Projecting this vector upon the normal to the illuminated element, and multiplying the length of this projection by the brightness, we obtain the flux density through this element.

2. Transformation of the Expression for the Solid Angle

If the brightness of a luminous surface S is equal to unity, the space illumination is measured by the solid angle subtended by S. The solid angle is

$$\omega = \int_S \frac{\cos(n, \mathbf{r})\, dS}{r^2} \tag{100}$$

where r is the radius vector from the apex of the solid angle to the element dS; and n is the direction of the normal to this element. Let us transform this surface integral into a contour integral. We describe a unit sphere about P (Fig. 30). The intersection of the cone with the sphere defines a spherical figure whose area is equal to the solid angle ω. The opposite half of the conical surface will cut from the sphere a symmetrical figure, the area of which is also equal to ω. Along an element of the contour, the tangents to the contour undergo a definite change in direction. Let us draw tangent planes to the conical surface, these planes being separated by an angle $d\gamma$, which characterizes the elementary change of direction of the tangent to the contour as we move along the contour. The planes intersect the sphere in two great circles

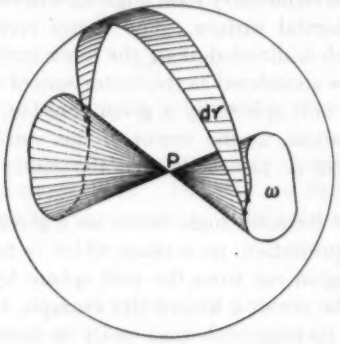


FIG. 30

and cut from it a lune, the area of which is equal to $2d\gamma$. Subdividing the contour into separate elements, and constructing at each of these points a plane that is tangent to the conical surface, we obtain an infinite number of elementary lunes formed by pairs of neighboring planes. These lunes cover the entire sphere, with the exception of the two areas cut from the sphere by the conical surface. In this manner

$$\oint 2\,d\gamma + 2\omega = 4\pi \tag{101}$$

or

$$\omega = 2\pi - \oint d\gamma \tag{102}$$

The integration is performed over the entire conical surface. For a

multi-sided solid angle, the integral becomes the sum of outer angles formed by the planes of this solid angle.

The author realizes that the foregoing simple method is not mathematically rigorous. The transformation could also be carried out in the classical manner but this procedure would be more laborious. In the manner we have treated the subject, the pyramid may be considered as a special case of a cone. The pyramid will cut from the sphere a spherical polygon, the number of sides of which we shall denote by m. To each side of this polygon corresponds an index i which varies from 1 to m. Let us denote by $\gamma_{i,i+1}$ the outer angle between the sides i and $i + 1$ (Fig. 31). With this notation, $\gamma_{m,m+1}$ corresponds to the

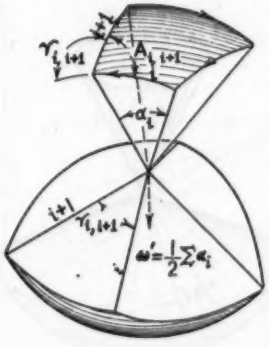


FIG. 31

angle $\gamma_{m,1}$ between the "last" and the "first" side of the polygon. The expression for the solid angle is

$$\omega = 2\pi - \sum_{i=1}^{m} \gamma_{i,i+1} \tag{103}$$

The amount to be added to the solid angle to obtain 2π is equal to the sum of the outer angles.

Denoting the angle between the inner sides of the pyramid by

$$A_{i,i+1} = \pi - \gamma_{i,i+1}$$

we may express the solid angle in the form

$$2\pi - \omega = \sum_{i=1}^{m} (\pi - A_{i,i+1}).$$

The amount to be added to the solid angle to obtain a hemisphere is equal to the sum of the supplements of the inner angles between the pyramid sides. For a three-sided angle ($m = 3$), we obtain as a special case the well known formula of spherical trigonometry,

$$\omega = \sum_{i=1}^{3} A_{i,i+1} - \pi \tag{104}$$

which states that the area of a spherical triangle is proportional to the spherical excess.

The following geometrical interpretation may be given to the contour integral expressing the solid angle. Consider a point within the conical surface and draw normals from it to the elements of the conical

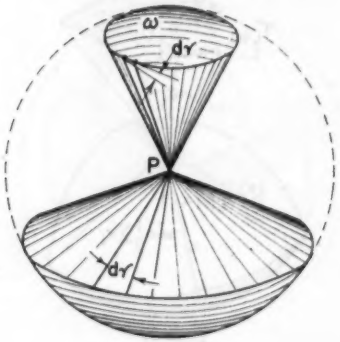


FIG. 32

surface. Drawing the outer normals as shown in Fig. 32, we obtain a new conical surface which may be called *polar* (generalizing the concept of polar triangles in spherical trigonometry) with respect to the original conical surface. To each tangential plane of the original cone corresponds a definite surface ray of the polar surface. The angle $d\gamma$ between two tangent planes is equal to the angle between the two normals to these planes, and these normals will be the two corresponding surface rays of the polar cone. Thus the contour integral of $d\gamma$ will be equal to the angle at the apex of the conical surface when the latter is cut along one of its rays and developed on a plane. The solid angle will be equal to the amount added to the apex angle to obtain 2π.

In the special case of a many-sided solid angle, the polar surface will

also be represented by a many-sided angle. To each side of the original pyramid will correspond an edge of the polar pyramid (Fig. 31). The outer angle between two planes forming the original pyramid will be equal to the plane angle between two adjacent edges of the polar pyramid. As known from geometry, the sum of plane angles composing a many-sided angle is less than 2π and is, as we have seen, less by the solid angle of the polar pyramid.

As an example, let us determine the solid angle subtended by the rectangle s when P is on a normal to the plane of s and the normal is drawn from one of the corners of s (Fig. 33). We number the sides of the rectangle and the corresponding sides of the pyramid, as shown in

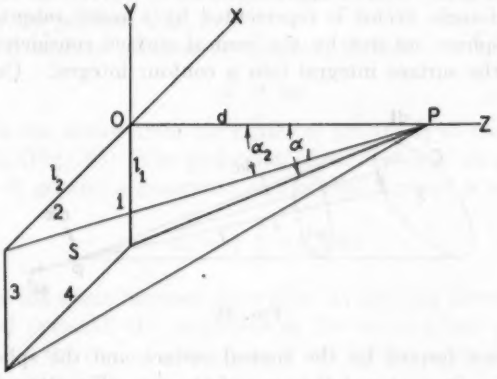


FIG. 33

the drawing. Three of the four angles, namely $\gamma_{1,2}$; $\gamma_{2,3}$; $\gamma_{4,1}$ are equal to $\pi/2$. Therefore,

$$\omega = \frac{\pi}{2} - \gamma_{3,4}$$

where ω is expressed in steradians and $\gamma_{3,4}$ is in radians. Since

$$\cos \gamma_{3,4} = \sin \alpha_1 \sin \alpha_2$$

we have

$$\omega = \sin^{-1} (\sin \alpha_1 \sin \alpha_2),$$

where α_1 and α_2 denote the plane angles subtended by the first and second sides of the rectangle. If α_1 and α_2 are small, $\omega \cong \alpha_1 \alpha_2$.

Starting from linear dimensions, it is convenient to use the following expression for the solid angle:

$$\omega = \tan^{-1} \frac{\frac{l_1}{\bar{d}} \cdot \frac{l_2}{\bar{d}}}{\sqrt{1 + \left(\frac{l_1}{\bar{d}}\right)^2 + \left(\frac{l_2}{\bar{d}}\right)^2}} \tag{105}$$

In conclusion, let us note that the formulas may also be useful in the computation of the flux through a given contour and originating from a uniform point source.

3. Transformation of the Expression for the Vector of the Solid Angle

The solid-angle vector is represented by a *vector integral* over the part of a sphere cut out by the conical surface considered. Let us transform the surface integral into a contour integral. Consider the

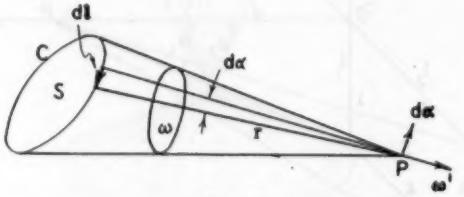


FIG. 34

closed surface formed by the conical surface and the sphere of unit radius having its center at the apex of the cone (Fig. 34). The vector of a closed surface is equal to zero. Indeed, the condition that a surface be closed is equivalent to the statement that the vectors of the elements of this surface form a closed figure. Thus the total vector obtained by the addition of the vector of the spherical surface and the vector of the conical surface is equal to zero, on the assumption that the vectors of separate elements are marked off along the outer normals to both the spherical and conical surfaces.

Subdivide the conical surface into its elements and consider the vector of the element of the conical surface enclosed between two rays which form the angle $d\alpha$. This element of the cone is a sector of a circle of unit radius with the angle $d\alpha$ at its apex, and its area is $d\alpha/2$. Each plane angle formed by two intersecting straight lines may be characterized by a vector directed along the normal to its plane and equal in length to the magnitude of the angle. Introducing the vector $d\boldsymbol{\alpha}$, we

find that the vector of the element of the conical surface is equal to $d\alpha/2$. Therefore the solid-angle vector is

$$\omega' = \tfrac{1}{2} \oint d\alpha \tag{106}$$

where the integration is performed over all the elements of the conical surface. When the base of the cone is a polygon and the cone degenerates into a pyramid, the solid-angle vector is

$$\omega' = \tfrac{1}{2} \sum_{i=1}^{m} \alpha_i \tag{106a}$$

where α_i is the vector of the plane angle α_i that is subtended by the i^{th} side of the polygon (Fig. 31).

It is easy to show (Section 7, Chapter IV) that

$$\omega' = r_0 \omega$$

where r_0 is the vector from the center of gravity Q to the apex of the solid angle (Fig. 20). The projection of the vector ω' on a direction n (which is in general a generatrix of a conical surface) is equal to

$$\omega'_n = \tfrac{1}{2} \oint \cos \beta \, d\alpha \tag{107}$$

where β is the angle between the vector $d\alpha$ and the direction n. For an m-sided pyramid, the projection of the vector ω' on a direction n is one half of the sum of projections of the vectors α_i on this direction. In other words,

$$\omega'_n = \tfrac{1}{2} \sum_{i=1}^{m} \alpha_i \cos \beta_i \tag{107a}$$

where β_i is the angle between the vector α_i and the direction n.

The above transformations are of great importance in theoretical photometry and they completely define the computation of the illumination produced by large uniformly luminous surfaces. The light vector of a luminous surface of brightness B is

$$\mathbf{D} = B\omega' = \frac{B}{2} \oint d\alpha. \tag{108}$$

Introducing the luminosity of the surface, we obtain $L = \pi B$, and

$$\mathbf{D} = L \oint \frac{d\alpha}{2\pi}. \tag{108a}$$

If the luminous surface is an m-sided polygon,

$$\mathbf{D} = L \sum_{i=1}^{m} \frac{\alpha_i}{2\pi} \tag{108b}$$

Thus the light vector produced by a luminous polygon is determined by constructing a pyramid with apex at P and with its base formed by the luminous polygon. We determine in degrees the angles $\alpha_1, \alpha_2 \cdots, \alpha_m$ in the sides of this pyramid, and consider vectors which are normal to the sides of the pyramid and whose lengths are equal to $\alpha_1/360$, $\alpha_2/360, \cdots, \alpha_m/360$. The sum of these vectors determines the light vector.

The flux density for any element of the surface is given by the projection of the light vector on the normal n to this element. Thus the flux density may be computed from the formula,

$$D_n = B\omega_n' = L \oint \frac{d\alpha}{2\pi} \cos\beta, \tag{108c}$$

For a polygon,

$$D_n = L \sum_{i=1}^{m} \frac{\alpha_i}{2\pi} \cos\beta_i, \tag{108d}$$

If the tangent plane does not intersect the luminous polygon, the quantity D_n represents the illumination. Thus the above formula gives a general expression for the illumination produced by an arbitrarily shaped luminous figure. These formulas make it unnecessary to perform an integration over a surface when considering special cases. Applied to the case of a luminous polygon, the computation of the illumination is reduced to the determination of the angles α_i and β_i.

Eq. (108d) was first derived by Lambert, who proceeded from the case of a spherical triangle, and was derived in a simpler manner by Wiener[22] and shown in its vectorial form for the first time by Mehmke.[21] For surfaces bounded by arbitrary contours, Eq. (108c) was derived in 1926 by Yamauti[29] and independently by the author of this book in collaboration with Gurevitch.[30] A somewhat different derivation is given in the paper by Genkin.[31] According to the principle of reci-

[29] Z. Yamauti, J. O. S. A., **13**, 1926, p. 561. Researches of the Electrotechnical Laboratory, Tokyo, No. 190 (in Japanese).

[30] A. Gershun and M. Gurevitch, The Light Field. Journal of the Russian Physico-Chemical Society, **60**, 1928, No. 4, p. 355.

[31] V. Genkin, Calcul de l'éclairement moyen en présence des surfaces diffusantes de brillance uniforme. R. G. E., **29**, 1931, p. 369.

procity, Eq. (108c) allows the determination of the flux through an arbitrary contour, the light being emitted according to Lambert's law by an element of the surface. The flux through this contour is equal to the product of the axial intensity of the luminous element and the projection of the vector of the solid angle on the axis. The computation is reduced to the determination of a contour integral or of a sum along the perimeter of a polygon.

4. **Example: Computation of Illumination from a Rectangular Source**

Let us determine the light vector at P produced by a uniformly luminous, perfectly diffusing rectangular source of brightness B_s' (Fig. 33). A table of cosines of the angles between the vectors α_i and the coordinate axes follows.

Vector	X	Y	Z
α_1	1	0	0
α_2	0	1	0
α_3	$-\cos \alpha_2$	0	$\sin \alpha_2$
α_4	0	$-\cos \alpha_1$	$\sin \alpha_1$

The vector

$$\omega' = \frac{1}{2} \sum_{i=1}^{4} \alpha_i$$

is determined from the table:

Vector	X	Y	Z
α_1	α_1	0	0
α_2	0	α_2	0
α_3	$-\alpha_3 \cos \alpha_2$	0	$\alpha_3 \sin \alpha_2$
α_4	0	$-\alpha_4 \cos \alpha_1$	$\alpha_4 \sin \alpha_1$
$\omega' = \frac{1}{2} \sum \alpha_i$	$\omega_x = \dfrac{\alpha_1 - \alpha_3 \cos \alpha_2}{2}$	$\omega_y = \dfrac{\alpha_2 - \alpha_4 \cos \alpha_1}{2}$	$\omega_z = \dfrac{\alpha_3 \sin \alpha_2 + \alpha_4 \sin \alpha_1}{2}$

From the diagram,

$$\tan \alpha_3 = \cos \alpha_2 \tan \alpha_1$$

and

$$\tan \alpha_4 = \cos \alpha_1 \tan \alpha_2 .$$

Thus,

$$D_x = B\omega_x' = \frac{B}{2}\{\alpha_1 - \cos\alpha_2 \tan^{-1}(\cos\alpha_2 \tan\alpha_1)\}$$

$$D_y = B\omega_y' = \frac{B}{2}\{\alpha_2 - \cos\alpha_1 \tan^{-1}(\cos\alpha_1 \tan\alpha_2)\} \tag{109}$$

$$D_z = B\omega_z' = \frac{B}{2}\{\sin\alpha_1 \tan^{-1}(\cos\alpha_1 \tan\alpha_2) + \sin\alpha_2 \tan^{-1}(\cos\alpha_2 \tan\alpha_1)\}$$

The absolute value of the light vector is

$$D = \sqrt{D_x^2 + D_y^2 + D_z^2} \tag{110}$$

where D_x is equal to the illumination at the point P in a plane parallel to the luminous rectangle, D_x is the illumination of the plane YZ, and D_y is the illumination of the plane XZ. When the illuminated plane is inclined but does not intersect the rectangle, its illumination is

$$E = D_x \cos(n, x) + D_y \cos(n, y) + D_z \cos(n, z). \tag{111}$$

If the linear dimensions of the sides of the rectangle are l_1 and l_2, and d is its distance from P, we obtain

$$\left.\begin{aligned} D_x &= \frac{B}{2}\left\{\tan^{-1}\left(\frac{l_1}{d}\right) - \frac{d}{\sqrt{d^2 + l_2^2}} \tan^{-1}\frac{l_1}{\sqrt{d^2 + l_2^2}}\right\} \\ D_y &= \frac{B}{2}\left\{\tan^{-1}\left(\frac{l_2}{d}\right) - \frac{d}{\sqrt{d^2 + l_1^2}} \tan^{-1}\frac{l_2}{\sqrt{d^2 + l_1^2}}\right\} \\ D_z &= \frac{B}{2}\left\{\frac{l_1}{\sqrt{d^2 + l_1^2}} \tan^{-1}\frac{l_2}{\sqrt{d^2 + l_1^2}} + \frac{l_2}{\sqrt{d^2 + l_2^2}} \tan^{-1}\frac{l_1}{\sqrt{d^2 + l_2^2}}\right\} \end{aligned}\right| \tag{112}$$

These formulas, it seems, were first derived by R. A. Herman.[32] The direction and the absolute value of the light vector from a rectangle are easily determined by graphical methods. We shall illustrate this, following the method of Mehmke[21] for daylight illumination through a window (Fig. 35). We draw from P the straight lines PQ_1, PQ_2, PQ_3, PQ_4, which enclose a region of the sky. As a rule this region will be bounded on its lower side by the plane passing through the roof of the opposite building. If the brightness B of the sky is assumed to be equal

[32] R. A. Herman, A Treatise on Geometrical Optics. Cambridge, 1900. Also see H. H. Higbie, Prediction of Daylight from Vertical Windows, I. E. S. Trans., 20, 1925.

H. H. Higbie and A. Levin, Prediction of Daylight from Sloping Windows, I. E. S. Trans., 21, 1926.

in all directions, and if loss of light in the glass is neglected, the relative value of the light vector **e** at P will be

$$\mathbf{e} = \frac{\mathbf{D}}{B\pi} = \sum_{i=1}^{4} \mathbf{e}_i, \tag{113}$$

where

$$\mathbf{e}_i = \frac{\mathbf{a}_i}{2\pi}. \tag{114}$$

The summation is performed over the four sides of the light pyramid which encloses the light rays arriving at P.

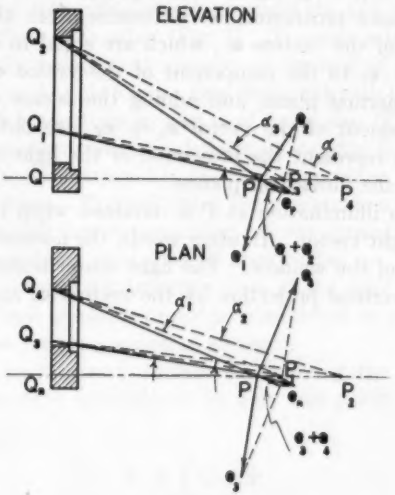


FIG. 35

For direct reading of the absolute values of the vectors $\mathbf{e}_i$ from the drawing, we may use a protractor suggested by Mehmke. An Archimedean spiral, i.e. a curve for which the radius vector is proportional to the angle, is drawn on tracing cloth. The radius vector is made of unit length for the angle 2π. To determine a length equal to $\alpha/2\pi$ we place the protractor on the drawing in such a manner that its center P coincides with the apex of the angle and the line PQ coincides with one of the sides of the angle. The spiral will cut from the other side of the angle the distance $\alpha/2\pi$, which is equal to the absolute value of the vector $\mathbf{e}_i$ corresponding to this angle α.

The solid angle has in this case the shape of a four-sided pyramid, two sides of which (1 and 2) are horizontal, while the other two sides (3 and 4) are vertical (Fig. 35). Thus the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ may be drawn and added on the vertical projection, and the vector $\mathbf{e}_3$ and $\mathbf{e}_4$ on the horizontal projection. To determine from the drawing the absolute values of these vectors, we imagine the planes of the angles α_1, α_2, α_3, α_4, to be placed parallel to the planes of the drawing. The angles α_1 and α_2 will be parallel to the vertical projection, and the angles α_3 and α_4 will be parallel to the plane of the horizontal projection. Then $P_1Q_n = PQ_1$; $P_2Q_n = PQ_2$; $P_3Q_p = PQ_3$; $P_4Q_p = PQ_4$ and $\angle Q_3P_1Q_4 = \alpha_1$; $\angle Q_3P_2Q_4 = \alpha_2$; $\angle Q_1P_3Q_2 = \alpha_3$; $\angle Q_1P_4Q_2 = \alpha_4$. Using the Mehmke protractor, we determine from these angles the absolute values of the vectors $\mathbf{e}_i$, which are equal to $\alpha_i/2\pi$. Adding the vector $\mathbf{e}_1 + \mathbf{e}_2$ to the component of the vector $\mathbf{e}_3 + \mathbf{e}_4$ that is parallel to the normal plane, and adding the vector $\mathbf{e}_3 + \mathbf{e}_4$ to the horizontal component of the vector $\mathbf{e}_1 + \mathbf{e}_2$, we obtain the vectors $\mathbf{e}_p$ and $\mathbf{e}_n$ which represent the projection of the light vector $\mathbf{e}$ on the vertical and on the horizontal planes.

The maximum illumination at P is obtained when its normal coincides with the light vector. In other words, the normal passes through the light center of the window. The light center is determined in the horizontal and vertical projection by the vectors $\mathbf{e}_p$ and $\mathbf{e}_n$.

CHAPTER VII

MATHEMATICAL THEORY

1. The Light Beam and Its Brightness

In geometrical optics, the light ray is considered as a mathematical straight line. In photometry (see Chapter II, Section 2) the light ray is replaced by an elementary ensemble of directions along which the transfer of radiant energy proceeds. The geometry of the light beam is defined by two infinitesimal apertures ds_1 and ds_2 which are at a finite distance from each other (Fig. 3). It seems advisable to introduce a new, purely geometrical concept—*a measure of the multitude of directions composing the light beam.* One deals, not with the infinitesimal apertures, but with apertures s_1 and s_2 of finite size (Fig. 36). Of all straight lines in space we choose a definite set passing through s_1 and s_2. All these lines are concentrated in the region τ, whose boundaries are shown by dotted lines. To each point P corresponds a definite conical surface, within which are concentrated all the straight lines passing through P and belonging to the set of lines within τ. To this conical surface, in turn, correspond definite values of the solid angle ω and of its vector $\boldsymbol{\omega}'$. These quantities may be considered as point functions; and outside of the region τ, ω and $\boldsymbol{\omega}'$ are zero.

We construct an arbitrary surface S satisfying the condition that each straight line of τ intersects it in only one point. Consider the quantity

$$N = \int_S \boldsymbol{\omega}' \cdot d\mathbf{S} \tag{115}$$

where the integration is performed over the entire surface S, and where $\boldsymbol{\omega}'$ is the vector of the solid angle enclosing all the lines of τ that pass through the element dS of surface S.

The quantity N represents the flux of $\boldsymbol{\omega}'$ through the surface S. N depends only on the size and relative position of the apertures and is independent of the choice of the surface S. The flux of the solid angle vector through all surfaces that satisfy the above requirement is the same. The proof of this statement follows from the fact that the vector $\boldsymbol{\omega}'$ is solenoidal (div $\boldsymbol{\omega}' = 0$).

Thus it seems natural to take as a *measure of the multitude of lines contained within* τ the quantity N which is an *invariant* with respect to all surfaces intersecting the set of lines. In place of S we may choose, as a special case, one of the apertures s_1 and s_2. If N is determined by two infinitesimal apertures, the measure of the multitude of the geometric rays composing the beam is

$$d^2N = d\sigma_1\, d\omega_1 = d\sigma_2\, d\omega_2 \tag{116}$$

where $d\sigma_1$ and $d\sigma_2$ are the areas of the cross-section of the beam at the two apertures and $d\omega_1$ and $d\omega_2$ are the corresponding values of the solid angle (Fig. 3). For other cross-sections of the light beam, we cannot give a definite value of the solid angle that measures the divergence of the rays, because this quantity will have different values for different points of such a cross-section, and it will decrease as we move from the center to the edge of the cross-section of the beam. Thus the common statement that "the product of the cross-sectional area of the beam and the solid angle determining divergence of the rays is constant" is not true. What is constant is the flux of the solid angle vector having for all the cross-sections the value

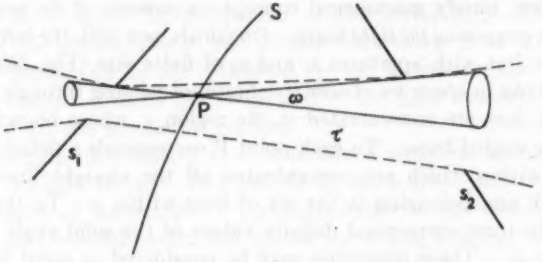


FIG. 36

$$d^2N = \frac{\cos\theta_1 \cos\theta_2\, ds_1\, ds_2}{l^2} \tag{117}$$

where l is the distance between the apertures ds_1 and ds_2, and θ_1 and θ_2 are the angles between the normals to the apertures and the direction of the beam.

The new concept of the solid-angle vector not only clarifies the somewhat foggy question of the geometry of the light beam, but may also

be used in building up a system of photometric concepts. Thus the brightness B along the light beam may be defined as the coefficient of proportionality between the light flux d^2F and the quantity d^2N, which measures the multitude of the geometrical rays composing the light beam:

$$d^2F = B\, d^2N. \tag{118}$$

Inasmuch as N is an invariant, the brightness varies along the beam in proportion to the flux. In empty space, the flux is constant for all the cross-sections of the beam and the brightness is also constant. Let us consider now the question of the change of brightness when the beam passes from one transparent medium into another. If the light beam passes from a medium (subscript k) in another medium (subscript

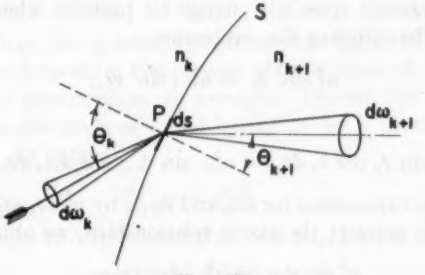


FIG. 37

$k + 1$, Fig. 37), the indices of refraction of these media being n_k and n_{k+1}, and if the boundary s of the two media is ideally polished, the change in direction of the beam at s is defined by the following:

$$n_k \sin \theta_k = n_{k+1} \sin \theta_{k+1}. \tag{119}$$

We shall now prove that $n^2\, d^2N$ is an invariant for all the points of the beam. *The invariability of this expression must be considered as one of the fundamental laws of geometrical optics.* The validity of this law for the special case of n = const was demonstrated in the foregoing discussion. We denote the area of the element of the refracting surface by ds and the point corresponding to it by P. The cross-sectional area of the light beam at this point will differ, depending upon the side from which we approach the refracting surface. If from the medium k, the area will be

$$d\sigma_k = ds \cos \theta_k.$$

If from the medium $k + 1$, it will be

$$d\sigma_{k+1} = ds \cos \theta_{k+1} .$$

Similarly, the solid angles defining the divergence of the light beam at P are denoted by $d\omega_k$ and by $d\omega_{k+1}$. The elementary solid angle $d\omega_k$ includes the directions for which the angle of incidence varies from θ_k to $\theta_k + d\theta_k$, and the position of the plane of incidence varies by an angle $d\psi_k$. Then

$$d\omega_k = \sin \theta_k \, d\theta_k \, d\psi_k \tag{120}$$

and

$$d\omega_{k+1} = \sin \theta_{k+1} \, d\theta_{k+1} \, d\psi_{k+1} . \tag{121}$$

We must consider that $d\psi_k = d\psi_{k+1}$ because the plane defined by the rays and the normal does not change its position when the rays are refracted. Differentiating the expression,

$$n_k^2 \sin^2 \theta_k = n_{k+1}^2 \sin^2 \theta_{k+1} \tag{122}$$

we obtain

$$n_k^2 \sin \theta_k \cos \theta_k \, d\theta_k = n_{k+1}^2 \sin \theta_{k+1} \cos \theta_{k+1} \, d\theta_{k+1} . \tag{123}$$

Multiplying the expressions for $d\omega_k$ and $d\omega_{k+1}$ by $n_k^2 \, d\sigma_k$ and by $n_{k+1}^2 \, d\sigma_{k+1}$ and taking into account the above relationships, we obtain

$$n_k^2 \, d\sigma_k \, d\omega_k = n_{k+1}^2 \, d\sigma_{k+1} \, d\omega_{k+1} . \tag{124}$$

For convenience of the proof we proceeded from the value of the quantity d^2N at the boundary point P. As has been shown, this quantity is the same for all points of the light beam in a given medium. From this it follows that the magnitude $n_k^2 \, d^2N_k$ is the same for all values of k. Along the entire path of the light beam, independent on how many times it has been refracted, the product of n^2 and the measure of the multitude of the rays remain constant. The invariability of the expression $n^2 \, d\sigma \, d\omega$ was first demonstrated by R. Straubel, and represents one of the most important laws of geometrical optics. Many papers deal with this question.[33]

[33] M. Herzberger, Über Sinusbedingung, Isoplanasie und Homöoplanasiebedingung, ihren Zusammengang mit energetischen Überlegungen und ihre Ableitung aus dem Fermatschen Gesetz. Zs. für Instrumentenkunde, **48**, 1928, pp. 313, 465, 524. (Contains a bibliography up to the year 1928.)

André Dargenton, Théorème sur la réfraction des pinceaux de rayons lumineux. Application au calcul de la brillance. Revue d'Optique, **8**, 1929, p. 4. Note

Analogous relationships may be written for the rays confined by the diaphragms s_1 and s_2 of finite size (Fig. 36). If the rays are refracted, the invariant is no longer represented by $N = \int_s \omega' \cdot d\mathbf{S}$, but by

$$n^2 \int_s \omega' \cdot d\mathbf{S}$$

where n is the index of refraction of the medium in which the surface S is placed. It is better to introduce as an invariant the quantity

$$\frac{1}{\lambda^2} \int_s \omega' \cdot d\mathbf{S}$$

where λ is the wavelength. This invariant represents a pure number which may play an important rôle in the study of the phenomena of diffraction.

We turn now from the questions of geometrical optics to the problems of photometry, and consider the change of brightness of a light beam as it passes from one medium to another. Denote the flux by d^2F_k for a point P_k in the medium k and by d^2F_{k+1} for a point P_{k+1} in the medium $k + 1$. In general,

$$d^2F_{k+1} < d^2F_k .$$

Denoting the values of brightness by B_k and B_{k+1}, we obtain

$$d^2F_k = B_k \, d^2N_k$$

and

$$d^2F_{k+1} = B_{k+1} \, d^2N_{k+1} .$$

Since $n_k^2 \, d^2N_k = n_{k+1}^2 \, d^2N_{k+1}$, we have

$$\frac{B_{k+1}}{n_{k+1}^2} : \frac{B_k}{n_k^2} = d^2F_{k+1} : d^2F_k .$$

The brightness along the light beam is directly proportional to the flux and to the square of the index of refraction, when passing from one medium into another. Disregarding the question of absorption of light, we obtain the invariability of the ratio B/n^2. From this

---

sur une propriété de la réfraction des pinceaux de rayons lumineux. Application au calcul de la brillance. Revue d'Optique, **12**, 1933, p. 172.

Pierre Copel, Sur un nouvel invariant relatif a l'ensemble de deux pinceaux lumineux ayant même rayon moyen. Rev. d'Optique, **13**, 1934, p. 193.

follows that the brightness of beams arriving at an image, produced by an ideal lens that does not absorb any light, is equal to the brightness of rays emitted by the object, provided that the image is in the same medium as the object.

The question of the change of brightness of a light beam when it is reflected from a boundary of two regions does not call for a separate analysis. The quantity d^2N remains unchanged in the process of reflection of a given light beam, and thus the ratio of the brightness of the reflected beam to the brightness of incident beam is equal to the ratio of the reflected flux to the incident flux. From this follows that the reflection factor of specular surfaces, which is equal by definition to the ratio of reflected flux to incident flux, may also be defined as the ratio of the brightness of the image to the brightness of the object. All these relationships are fundamental in the study of photometry of the optical image.[34]

2. Some Mathematical Relationships

Let us consider some of the most important theorems of the mathematical theory of the light field in empty space. We shall treat them briefly, inasmuch as the purpose of this text is to present the physical basis of the theory of the light field and its application in radiation engineering. The mathematical theory of the light field is discussed in papers by Fock,[35] Gershun,[19] Gurevitch[30] and others. A review of the literature appears in a paper by Yamauti.[7]

At first sight it would seem that the theory of the light field in its systematic treatment must be very like the theory of the electrostatic field. The light vector produced by a point source varies inversely as the square of the distance, as does the electric force produced by a point charge. With a linear source, the light vector varies inversely as the first power of the distance, as in the electric field of a charged wire. With an infinite plane source, the light vector is constant in direction and magnitude at any finite distance from the plane, as in the homogeneous field of a charged plane. However, the light field possesses a

[34] A. I. Tudorovsky, Elements of the General Theory of Optical Apparatus. Edited by "The Military Technical Academy" PKKA of Dzerzinsky, 1932. M. M. Gurevitch, Light Measurements. Edited by the Educational Dept. of the Civil Air Fleet, Leningrad, 1934.

[35] V. A. Fock, Illumination Produced by Surfaces of Arbitrary Shape. Proc. of State Opt. Inst., **3**, No. 28, 1924.

V. Fock, Zur Berechnung der Beleuchtungsstärke. Zs. f. Phys., **38**, 1924, p. 102.

number of curious properties, the study of which may be of general interest from the standpoint of field theory.

The electric charge, or the strength of a magnetic pole, does not depend on the direction in which its action is considered. The value of the force vector is completely defined by the distance from the source. The field produced by uniform point sources of light, the intensity of light of which is the same in all directions, is identical with the field produced by point charges. In this special case, the light vector possesses a scalar potential. It is possible that the historical error of Euler and Laplace, who assumed that the luminous intensity of each element of a luminous surface is the same in all directions, was caused by a subconscious desire to introduce an analogy to the theory of gravitation, where the mass does not depend on direction. This error was discovered by Lambert, who assumed that the luminous intensity varies according to the cosine law.

Usually the luminous intensity of a point source is different in different directions, and the value of the light vector depends on both distance from the source and direction. Thus electrostatic and mechanical analogs cannot be used. The light field usually has vortices and the light vector does not possess a potential. In fields produced by several sources, in general, a potential does not exist and it is impossible to construct surfaces orthogonal to the flux lines.

Let us consider now some vectorial operations and relationships characterizing the light field in empty space. First, let us introduce the concept of the vector potential. If the light vector **D** is solenoidal (div **D** = 0), the flux of this vector through a surface s may be expressed, according to Stokes' theorem, as a line-integral of another vector **A** over a contour C embracing this surface, or

$$\int_s \mathbf{D} \cdot d\mathbf{s} = \oint \mathbf{A} \cdot d\mathbf{l}.$$

The vector **D** is the curl of the vector **A**, which is called the *vector potential*, or

$$\mathbf{D} = \text{curl } \mathbf{A}. \tag{125}$$

The field of the vector **A** defines the field of the light vector **D**. The value of the concept of vector potential in photometry resides in the fact that it allows a simplified computation of flux from one surface to another. The integral of the light vector over a surface is reduced to the integral of the vector potential along a contour. Let us recall,

first, the well-known relationships between the components of the vectors **D** and **A** in rectangular system of coordinates, or

$$\left.\begin{aligned} D_x &= \operatorname{curl}_x \mathbf{A} = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \\ D_y &= \operatorname{curl}_y \mathbf{A} = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \\ D_z &= \operatorname{curl}_z \mathbf{A} = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \end{aligned}\right\} \tag{126}$$

These relationships determine the components D_x, D_y, D_z of curl **A** when the values A_x, A_y, A_z are known at all points of the field.

We consider now the inverse problem of finding the vector potential. For a uniformly luminous surface which radiates according to Lambert's law, the light vector is

$$\mathbf{D} = \frac{B}{2} \oint d\mathbf{a} \tag{127}$$

Denoting by r the distance from the element dl to the point P, and introducing a unit vector $\mathbf{r}_1$,

$$d\mathbf{a} = \frac{\mathbf{r}_1}{r} \times d\mathbf{l} \tag{128}$$

The components of the vector $\frac{\mathbf{r}_1}{r}$ along the coordinate axis may be considered as derivatives of a scalar function $\ln r$ along the corresponding axis, or

$$\frac{\mathbf{r}_1}{r} = \operatorname{grad} \ln r \tag{129}$$

Therefore,

$$d\mathbf{a} = \operatorname{grad} \ln r \times d\mathbf{l} = \operatorname{curl} (\ln r\, d\mathbf{l}) \tag{130}$$

The identity of the two above expressions for $d\mathbf{a}$ may easily be verified by substitution.

Thus the solid-angle vector is

$$\omega' = \frac{1}{2} \oint d\mathbf{a} = \frac{1}{2} \oint \operatorname{curl} (\ln r\, d\mathbf{l}) = \operatorname{curl} \frac{1}{2} \oint \ln r\, d\mathbf{l} \tag{131}$$

The light vector is

$$\mathbf{D} = B\omega' = \text{curl } \mathbf{A}$$

and its vector potential $\mathbf{A}$ is

$$\mathbf{A} = \frac{B}{2} \oint \ln r \, d\mathbf{l} \tag{132}$$

The luminous flux F from the luminous surface S of brightness B (enclosed by the contour C) to the illuminated surface s (enclosed by the contour c) is

$$\begin{aligned} F &= \int_s \mathbf{D} \cdot d\mathbf{s} = \int_s \mathbf{A} \cdot d\mathbf{l}' = \frac{B}{2} \int_s \int_c \ln r \, d\mathbf{l} \cdot d\mathbf{l}' \\ &= \frac{B}{2} \int_s \int_c \ln r \cos (dl, dl') \, dl \, dl' \end{aligned} \tag{133}$$

where r is the distance between dl and dl'. The computation of the flux is reduced to a two-fold integration in place of a four-fold integration. From the principle of reciprocity, the expression for F is symmetrical with respect to the elements of the contours of the luminous as well as of the illuminated surfaces.

Photometric computation using this formula has been carried out. V. Fock[35] computed the illumination produced by a luminous ellipse, using the foregoing method, while V. Genkin[31] considered the computation of the flux from one rectangle to another. The concept of the vector potential was introduced into photometric theory by R. Herman.[32]

To conclude, let us consider in greater detail the question of the existence of a scalar potential. Each luminous body may be subdivided into elements which may be considered as point sources. Thus we shall start with a light field produced by a point source which has an arbitrary intensity distribution. Let us denote by r the distance from the source to a point of the field, by $\mathbf{r}_1$ a unit vector, and by I the intensity of the source in the direction of $\mathbf{r}_1$. Then the light vector is

$$\mathbf{D} = \frac{I}{r^2} \mathbf{r}_1$$

The divergence is $\text{div } \mathbf{D} = \dfrac{\text{grad } I \cdot \mathbf{r}_1}{r^2}$. Inasmuch as *grad* I is normal to $\mathbf{r}_1$,

$$\text{div } \mathbf{D} = 0. \tag{134}$$

This relationship will be true for the light vector produced by any set of such sources and will therefore apply for an arbitrary luminous body. *The light vector is solenoidal.* This property was also found in Section 6, Chapter IV, on the basis of the law of conservation of energy.

The curl of the light vector is

$$\operatorname{curl} \mathbf{D} = \frac{\operatorname{grad} I \times \mathbf{r}_1}{r^2} \tag{135}$$

The curl is normal to $\mathbf{r}_1$ and to *grad* I. Inasmuch as these two vectors are in their turn normal to each other, the absolute value of the curl is

$$|\operatorname{curl} \mathbf{D}| = \left| \frac{\operatorname{grad} I}{r^2} \right| \tag{136}$$

Curl **D** is different from zero, and therefore the field of a point source does not, in general, possess a potential. The curl of the light vector is normal to that vector, so

$$\mathbf{D} \cdot \operatorname{curl} \mathbf{D} = 0. \tag{137}$$

This relationship reveals the existence of surfaces orthogonal to **D**. The flux lines represent the geodesics of the function $\phi = 1/r$ which may be called the *quasi-potential*, because

$$\mathbf{D} = -I \operatorname{grad} \phi. \tag{138}$$

When the light vector is produced by a set of non-uniform point sources,

$$\mathbf{D} \cdot \operatorname{curl} \mathbf{D} \neq 0$$

so the light vector possesses neither a potential nor a quasi-potential. It is impossible in this case to construct surfaces orthogonal to the flux lines. The exceptions to this rule are the two-dimensional field, the field possessing axial symmetry, the field produced by a luminous plane source that is perfectly diffusing.

In all these cases, the light vector may be represented by two variable scalar functions:

$$\mathbf{D} = -\psi \operatorname{grad} \phi \tag{139}$$

and the light lines may be constructed as orthogonal trajectories to surfaces upon which the values of the quasi-potential ϕ are constant. An example is the field produced by a plane, perfectly diffusing source.

Then ψ represents the distance along the normal from the luminous surface S to an arbitrary point of the light field, and

$$\phi = \frac{1}{2}\int_s \frac{B\,dS}{r^2} \tag{140}$$

where r is the distance from the point to the element dS whose brightness B is equal in all directions (A. Gershun. See footnote 19). Of unquestionable interest from the physical as well as from the engineering standpoint is the theory of the light field in diffusing, absorbing, and luminous media.[36] Space forbids a consideration of this question and a large number of other photometric problems.

[36] N. Boldyreff, see footnote 18.

A. Gershun, On Photometry of Diffusing Media. Proc. of State Opt. Inst., 1936.

A. Gershun, Photometric Computation of Gas Discharge Lamps. Svetotechnika, 1932, No. 8–9, p. 10.

A. Gershun, Berechnung des Volumleuchtens. Phys. Zs. d. Sowjetunion, **2**, No. 2, 1932.

A FUNCTION RELATED TO THE SERIES FOR e^{e^x}

By Leo F. Epstein

I. **Introduction and Definitions.** In deriving a series expansion for e^{e^x}, an interesting group of integral constants is encountered, the coefficients in the expansion

$$(1) \qquad e^{e^x} = e\left[1 + K_1 x + \frac{K_2}{2!}x^2 + \frac{K_3}{3!}x^3 + \cdots + \frac{K_n}{n!}x^n + \cdots\right]$$

The numerical values of these constants can be obtained most directly by substituting the value $x = 0$ in the derivatives of e^{e^x}, the n'th derivative thus yielding K_n, following the rule for a Maclaurin[1] series. In this paper, we shall discuss a few of the simpler properties of these integers, and of a function derived from them.

As a study of the derivation from the derivatives of e^{e^x} will show, the K-numbers (as we shall call them) are given by the following iteration formula

$$(2) \qquad K_{n+1} = 1 + nK_1 + \frac{n(n-1)}{2!}K_2 + \frac{n(n-1)(n-2)}{3!}K_3 + \cdots \cdots + nK_{n-1} + K_n$$

$$K_1 = 1.$$

$$(3) \qquad K_{n+1} = 1 + \sum_{m=1}^{n}\binom{n}{m}K_m$$

where $\binom{n}{m}$ is the ordinary binomial coefficient. Thus K_n can be calculated at once if the values of K_1 to K_{n-1} are known following the above equations. In this respect, the K-numbers resemble Euler's numbers, which are also integers derivable from similar equations involving binomial coefficients providing all the values from E_1 to E_{n-1} are known. Since the binomial coefficients, which are the multiples of the K's,

[1] Whittaker and Watson *Modern Analysis* (Cambridge, 1935).

COLE MEMORIAL LIBRARY

are all whole numbers, and since $K_1 = 1$, K_n *must always be an integer.* It should be observed that symbolically, we may write[2]

$$K_{n+1} = (1 + k)^n \tag{4}$$

where, in the expanded series, we substitute K_m for k^m in each term. It has further been demonstrated by various writers[3] that we may express K_n in the infinite series form

$$K_n = \frac{1}{e} \sum_{t=0}^{\infty} \frac{t^n}{t!} \qquad (n = 1, 2, 3, \cdots) \tag{5}$$

This equation is restricted to positive integral values of n; we shall now demonstrate that this restriction is unnecessary. In fact, we shall define the K-function

$$K_z = \frac{1}{e} \sum_{t=0}^{\infty} \frac{t^z}{t!} \tag{6}$$

where z is any number, real or complex. This definition is quite valid, for (6) converges absolutely for all values of z: applying the well-known ratio test, the ratio of the $(t + 1)$st term to the t'th term is

$$\frac{(t+1)^z}{(t+1)!} \cdot \frac{t!}{t^z}$$

and as t becomes infinite, the limit of this expression is identically zero for all finite values of z, real or complex.

By simple algebraic manipulation of the definition of K_z (Equation 6), we obtain, when x is a real positive integer,

$$eK_x = \frac{x^x}{x!} + (x-1)^x \sum_{n=1}^{\infty} \left[\frac{1}{(x-n)!} \left(\frac{x-n}{x-1} \right)^x + \frac{1}{(x+n)!} \left(\frac{x+n}{x-1} \right)^x \right]$$

the first term of the sum vanishing for $n \geqq x$. This equation is valid for all values of $x \geqq 0$, and is especially useful for the calculation of K_x for very large values of the argument.

[2] Equation (4) was known to d'Ocagne: Am. Jour. Math., 9, 370 (1887).

[3] This equation (5) for positive integers, the coefficients in the series expansion for e^{e^x} was given by Dobiński: Grunert's Archiv., 61, 333 (1877), and is said to have been known by Euler (cf. Bell, footnote 24c). Another early reference to these numbers, which seem to have been rediscovered many times is Cesaro: Nouvelles Annales de Math., 4, 39 (1885).

For negative values of z, equation (6) assumes the Dirichlet series form

$$K_{-z} = \frac{1}{e} \sum_{t=0}^{\infty} \frac{1}{t^z \cdot t!} \tag{7}$$

Although equations (6) and (7) both converge absolutely for all values of z, it will be observed that while (7) converges very rapidly, the

TABLE I

K-Numbers for Positive and Negative Integers

n	K_n	K_{-n}
0	$1 - 1/e$	$1 - 1/e = 0.63212\ 05588$
1	1	.48482 91072
2	2	.42177 34383
3	5	.39340 93945
4	15	.38019 78350
5	52	.37389 58961
6	203	.37084 15557
7	877	.36934 54823
8	4,140	.36860 75427
9	21,147	.36824 18738
10	115,975	.36806 01230
11	678,570	.36796 96382
12	4,213,597	.36792 44638
13	27,644,437	.36790 19333
14	190,899,322	.36789 06810
15	1,382,958,545	.36788 50591
16	10,480,142,147	.36788 22496
17	82,864,869,804	.36788 08447
18	682,076,806,159	.36788 01431
19	5,832,742,205,057	.36787 97922
20	51,724,158,235,372	.36787 96167
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
∞	∞	$1/e =$ .36787 94412

successive terms continually becoming smaller for all positive values of z, and approaching a limiting value most rapidly when z is large, the lower terms (small t) of equation (6) actually diverge at first since each term is greater than the preceding one, but eventually, when $t! > t^z$, the terms become successively smaller and the series converges. This initial divergence of the terms of (6) becomes increasingly apparent as we attempt to calculate K_z for large positive values of x from it. An asymptotic expansion, valid for large values of z is consequently highly desirable, and we shall show later how such an expansion can be obtained. For the exact computation of K_n equation (3) is most convenient; Table I has been calculated[4] using these principles.

We may derive several interesting relations from equation (7). Thus when $z = 1$

$$K_{-1} = \frac{1}{e}\sum_{t=1}^{\infty}\frac{1}{t\cdot t!}$$

Now since

$$\frac{e^x}{x} = \frac{1}{x} + \sum_{t=1}^{\infty}\frac{x^{t-1}}{t!},$$

$$\int_0^x \frac{e^x}{x}\,dx = \ln x + \sum_{t=1}^{\infty}\frac{x^t}{t\cdot t!}$$

If $x = 1$

$$\int_0^1 \frac{e^x}{x}\,dx = \ln 1 + \sum_{t=1}^{\infty}\frac{1}{t\cdot t!},$$

or

$$K_{-1} = \frac{1}{e}\int_0^1 \frac{e^x}{x}\,dx \tag{8}$$

[4] For this table, and other numerical computations, free use was made of the following tables:

a. Fry *Probability and its Engineering Uses*. New York, 1928, Appendix III.

b. Hayashi *Sieben- und Mehrstellige Tafeln*. Berlin, 1926.

c. *Barlow's Tables* (ed. by Comrie). London, 1935.

d. Oakes *Tables of Reciprocals*. London, 1865.

e. Davis *Tables of the Higher Mathematical Functions*, v. I and II. Bloomington, Ind., 1933 and 1935.

f. Thompson *Logaritmetica Brittanica*. Cambridge U. Press, 1924–1937.

g. *Smithsonian Mathematical Formulae* (Adams, ed.), Washington, 1922.

A *Millionaire* computing machine with a ten place keyboard was employed. The values of K_{12}, K_{13}, and K_{20} were not computed independently, but were taken from the paper by Bell (cf. footnote 24c).

The exponential integral $\int_0^t \frac{e^x}{x}\,dx$ can be found tabulated in various places,[5] and can be used to calculate K_{-1} very accurately. Similarly we can readily prove

$$K_{-2} = \frac{1}{e}\int_0^1 \frac{1}{x_1}\int_0^{x_1} \frac{e^{x_2}}{x_2}\,dx_2\,dx_1$$

$$K_{-3} = \frac{1}{e}\int_0^1 \frac{1}{x_1}\int_0^{x_1} \frac{1}{x_2}\int_0^{x_2} \frac{e^{x_3}}{x_3}\,dx_3\,dx_2\,dx_1$$

or in general

$$(9)\quad K_{-n} = \frac{1}{e}\int_0^1 \frac{1}{x_1}\int_0^{x_1} \frac{1}{x_2}\int_0^{x_2} \frac{1}{x_3}\cdots\frac{1}{x_{n-1}}\int_0^{x_{n-1}} \frac{e^{x_n}}{x_n}\,dx_n\cdots dx_2\,dx_1$$

a continued integral form, made up of n integrations, the final one with the limits zero to one. This equation applies, of course, only to integral values of n.

II. **Properties of the K-Function.** In this section we shall develop equations for some of the properties of the K-function. First, by Taylor's expansion

$$K_{x+\xi} = K_x + \xi K_x' + \frac{\xi^2}{2!}K_x'' + \cdots + \frac{\xi^m}{m!}K_x^{(m)} + \cdots$$

$$(10)\quad K_x^{(m)} = \frac{1}{e}\sum_{t=1}^{\infty}\frac{t^x}{t!}(\ln t)^m$$

so that

$$(11)\quad K_{x+\xi} = \frac{1}{e}\sum_{m=0}^{\infty}\frac{\xi^m}{m!}\sum_{t=1}^{\infty}\frac{t^x}{t!}(\ln t)^m$$

This equation is most useful for computing K_y for large non-integral values of y, taking $y = x + \xi$, where ξ is less than one. The series converges for all values of x and ξ; for the coefficient of $\frac{\xi^m}{m!}$ in each case is

$$L_m(x) = \sum_{t=1}^{\infty}\frac{t^x}{t!}(\ln t)^m$$

---

[5] a) Jahnke-Emde *Funktionentafeln*. Leipzig, 1933.
b) British Assoc. Adv. Sci., Mathematical Tables I. London, 1931.

The ratio of the $(t+1)$st to the t'th term of this series is

$$\left[\frac{\ln (t+1)}{\ln t}\right]^m \left(1+\frac{1}{t}\right)^x \frac{1}{t+1}$$

and as t approaches infinity, this approaches

$$\lim_{t\to\infty}\left\{\frac{1}{t+1}\left[\frac{\ln t\left(1+\frac{1}{t}\right)}{\ln t}\right]^m\right\} = 0, \text{ for all values of } m, \text{ finite or infinite.}$$

The series for $L_m(x)$ thus converges to a constant value. The ratio of the $(m+1)$st to the m'th term of equation (14) is

$$\frac{\xi^{m+1}}{\xi^m}\cdot\frac{m!}{(m+1)!}\cdot\frac{L_{m+1}}{L_m} = \frac{\xi}{m+1}\cdot\frac{L_{m+1}}{L_m}$$

Then as $m \to \infty$, this equation also approaches zero as a limit, since the ratio of the two finite constants $\frac{L_{m+1}}{L_m}$ is always finite. Thus equation (14) converges for all values such that

$$-\infty < \xi < +\infty.$$

As a special case of (14), we have the Maclaurin series expansion for K_x

$$K_x = \frac{1}{e}\sum_{m=0}^{\infty}\frac{p_m}{m!}x^m \tag{12}$$

the coefficients of this series

$$p_m = \sum_{t=1}^{\infty}\frac{(\ln t)^m}{t!}$$

have been calculated, and are summarized in Table II. A check on this computation is obtained from the interesting relation

$$\sum_{m=1}^{\infty}\frac{p_m}{m!} = \sum_{m=1}^{\infty}\frac{1}{m!}\sum_{t=1}^{\infty}\frac{(\ln t)^m}{t!} = 1$$

which can be derived from (12). The power series formulation is limited in its usefulness for calculating K_x only by the wearily slow convergence of the series for large values of x.

TABLE II

Coefficients of the Power Series Expansion for K_x

m	p_m	$p_m/m!$	$\sum_{k=1}^{m} p_k/k!$
0	1.71828 18285	—	—
1	0.60378 28628	0.60378 28628	.60378 28628
2	.54837 82849	.27418 91425	.87797 20053
3	.54296 35131	.09049 39188 7	.96846 59241
4	.58570 49319	.02440 43721 6	.99287 02963
5	.68236 68995	.00568 63908 29	.99855 66871
6	.84815 49252	.00117 79929 52	.99973 46801
7	1.11172 0107	.00022 05793 851	.99995 52594
8	1.52259 0134	.00003 77626 5211	.99999 30221
9	2.16436 2125	.00000 59644 01799	.99999 89865
10	3.17790 0086	.00000 08757 44071	.99999 98622
11	4.80224 7807	.00000 01203 06433	.99999 99826
12	7.44736 1338	.00000 00155 47675	.99999 99981
13	11.82463 354	.00000 00018 98923	1.00000 00000
14	19.18356 156	.00000 00002 20050	1.00000 00002
15	31.74519 554	.00000 00000 24276	1.00000 00002

The last figure is uncertain in the second and fourth columns.

We have already considered the differentiation of K_x in equation (10). Integration formulae may be readily derived from our definition (6). Thus

$$\int K_x f(x)\,dx = \frac{1}{e}\sum_{t=0}^{\infty}\frac{1}{t!}\int f(x)e^{x^{\ln t}}\,dx \tag{13}$$

A large number of indefinite and definite integrals involving K_x may be evaluated using this equation.

So far, we have discussed K_x only as a function of a real variable; if $x = iy$

$$K_{iy} = \frac{1}{e}\sum_{t=0}^{\infty}\frac{t^{iy}}{t!} = \frac{1}{e}\left[\sum_{t=0}^{\infty}\frac{\cos(\ln t)y}{t!} + i\sum_{t=0}^{\infty}\frac{\sin(\ln t)y}{t!}\right] \tag{14}$$

that is, K_{iy} is a complex quantity. Similarly, for a complex argument

$$K_{x+iy} = \frac{1}{e}\left[\sum_{t=0}^{\infty}\frac{t^x\cos(\ln t)y}{t!} + i\sum_{t=0}^{\infty}\frac{t^x\sin(\ln t)y}{t!}\right] \tag{15}$$

In this last equation, we have terms of the form

$$\frac{t^x \begin{matrix}\sin\\ \cos\end{matrix} (\ln t)y}{t!}$$

since neither $\sin x$ nor $\cos x$ is greater than one

$$\left| \frac{t^x \begin{matrix}\sin\\ \cos\end{matrix} (\ln t)y}{t!} \right| \leqq \frac{t^x}{t!}$$

But $\frac{t^x}{t!}$ is the t'th term of the series for K_x which we have already proven absolutely convergent. Thus following the comparison test for convergence,[6] equation (15) is absolutely convergent for all values of x and y; and therefore, equation (14) which is a special case of (15) is also absolutely convergent. Now since $e^{x+iy} = e^{x+i(y+2\pi k)}$

$$K_{x+i(y+2\pi k)} = K_{x+iy} \tag{16}$$

where k is an integer, and it follows that the K-function is periodic with the imaginary period $2\pi i$.

We shall conclude this section with a number of simply derived summation formulas, which we write without proof:

$$\sum_{n=0}^{\infty} \frac{K_n}{n!} = \frac{1}{e}(e^e - 1) = 5.2056 \cdots \tag{17}$$

$$\sum_{n=0}^{\infty} (-1)^n \frac{K_n}{n!} = \frac{1}{e}(e^{1/e} - 1) = 0.16359 \cdots$$

$$\begin{cases} \displaystyle\sum_{n=0}^{\infty} \frac{K_{2n}}{(2n)!} = \frac{1}{2e}(e^e + e^{1/e} - 2) \\ \displaystyle\sum_{n=0}^{\infty} \frac{K_{2n+1}}{(2n+1)!} = \frac{1}{2e}(e^e - e^{1/e}) \end{cases} \tag{18}$$

$$\begin{cases} e^{e^x \cos y} \cos(\sin y) = e + e \displaystyle\sum_{n=1}^{\infty} \frac{K_n}{n!} r^n \cos n\theta \\ e^{e^x \cos y} \sin(\sin y) = e \displaystyle\sum_{n=1}^{\infty} \frac{K_n}{n!} r^n \sin n\theta \end{cases} \tag{19}$$

In these last interesting Fourier series, the symbols r and θ have their usual polar-coordinate significance.

[6] Cf. Whittaker and Watson (1), ¶2.34.

III. **An Asymptotic Expansion**[7] **for K_n.** In this section, we shall attempt to develop a more convenient asymptotic formula for the function considered above. Briefly, the procedure to be followed consists of two steps: (A) the convergent infinite series formulation of K_n (equation 6) is converted into an asymptotically equivalent infinite integral by the use of the Euler-Maclaurin sum formula; and (B) this integral is then approximated asymptotically using a method developed by Laplace.

(A) We start then with the Euler-Maclaurin equation[8]

$$\begin{aligned}\sum_{x=0}^{\infty} f(x) = \int_0^{\infty} f(x)\,dx - \frac{1}{2}[f(\infty) - f(0)] + \frac{B_1}{2!}[f'(\infty) - f'(0)] \\ - \frac{B_2}{4!}[f'''(\infty) - f'''(0)] + \cdots \\ + (-1)^m \frac{B_m}{(2m)!}[f^{(2m-1)}(\infty) - f^{(2m-1)}(0)] \\ + \cdots\end{aligned} \tag{20}$$

In the case of the K-function

$$f(x) = \frac{x^n}{\Gamma(x+1)} \tag{21}$$

where the Gamma function now replaces the factorial in the denominator in order to obtain a continuous analytic function. Then we may write

$$\begin{aligned}eK_n = \int_0^{\infty} \frac{x^n\,dn}{\Gamma(x+1)} - \tfrac{1}{2}[f(\infty) - f(0)] + \tfrac{1}{12}[f'(\infty) - f'(0)] \\ - \frac{1}{720}[f'''(\infty) - f'''(0)] + \frac{1}{30,240}[f^v(\infty) - f^v(0)] \\ - \cdots \\ \cdots - (-1)^m \frac{B_m}{(2m)!}[f^{(2m-1)}(\infty) - f^{(2m-1)}(0)] \\ + \cdots\end{aligned} \tag{22}$$

[7] The restriction to integral values of the argument is made here only for convenience in formulation. Many of the results finally obtained, it will subsequently appear, are also valid for K_z, where z is non-integral.

[8] Cf. Ford *Studies on Divergent Series*, Macmillan (1916); also Ford *Asymptotic Developments* (U. of Mich., 1936).

or since $f(\infty)$ and $f(0)$ are both zero, we may write

$$eK_n = \int_0^\infty \frac{x^n\,dx}{\Gamma(x+1)} - s_n \tag{23a}$$

where

$$s_n = \sum_{m=1} (-1)^m \frac{B_m}{(2m)!}\left[f^{(2m-1)}(\infty) - f^{(2m-1)}(0)\right] \tag{23b}$$

We shall now evaluate $f^{(k)}(\infty)$ and $f^{(k)}(0)$ using Leibnitz's theorem[9] in the form

$$f^{(k)}(x) = \sum_{i=0}^{k} \binom{k}{i} \frac{d^i}{dx^i}(x^n)\,\frac{d^{k-i}}{dx^{k-i}}\left[\frac{1}{\Gamma(x+1)}\right] \tag{24}$$

But since

$$\frac{d^i}{dx^i}(x^n) = n(n-1)(n-2)\cdots(n-i+1)x^{n-i}$$

$$f^{(k)}(x) = \sum_{i=0}^{k} \binom{k}{i}\{n(n-1)(n-2)\cdots(n-i+1)x^{n-i}\} \cdot \frac{d^{k-i}}{dx^{k-i}}\left[\frac{1}{\Gamma(x+1)}\right] \tag{25}$$

We shall first use (25) to prove $f^{(k)}(\infty) = 0$; since $1/\Gamma(x)$ resembles e^{-x} in its behavior at $x = \infty$, in that the function and all of its finite derivatives are zeros of infinite order, it follows that

$$f^{(k)}(\infty) = \operatorname*{Lim}_{x\to\infty} \frac{d^m}{dx^m}\left[\frac{x^n}{\Gamma(x+1)}\right] = 0 \tag{26}$$

Again from (25), we observe that $f^{(k)}(0)$ is zero unless $i = n$. When this is true

$$f^{(k)}(0) = \binom{k}{n} \cdot n!\,\frac{d^{k-n}}{dx^{k-n}}\left[\frac{1}{\Gamma(x+1)}\right]_{x=0}$$

Such a term will occur only when $k \geqq n$; so that

$$\begin{aligned} f^{(k)}(0) &= 0 && \text{when } 0 < k < n \\ &= \frac{k!}{(k-n)!}\,\frac{d^{k-n}}{dx^{k-n}}\left[\frac{1}{\Gamma(x+1)}\right]_{x=0} && \text{when } k \geqq n \end{aligned} \tag{27}$$

[9] Cf. e.g. Hardy *Pure Mathematics*, Cambridge (1933).

Now Weirstrass long ago demonstrated that the reciprocal of the Gamma function can be developed in a convergent power series

$$\frac{1}{\Gamma(x)} = \sum_{k=1}^{\infty} a_k x^k$$

or

$$\frac{1}{\Gamma(x+1)} = \sum_{k=1}^{\infty} a_k x^{k-1}$$

From Taylor's theorem, we observe that the expression required for (27) can be written most simply as

$$\frac{d^{k-n}}{dx^{k-n}}\left[\frac{1}{\Gamma(x+1)}\right]_{x=0} = a_{k-n+1}\cdot(k-n)!$$

Whence

$$f^{(k)}(0) = 0 \qquad (0 < k < n) \tag{28}$$

$$= k!\cdot a_{k-n+1} \qquad (k \geqq n)$$

Substituting the values of the derivatives at infinity and at the origin (equations 26 and 28) in (23b), we obtain

$$s_n = -\sum_{m=m_0} (-1)^m \frac{B_m}{2m}\cdot a_{2m-n} \tag{29}$$

where

$$m_0 = \begin{cases} (n+1)/2 & \text{when } n \text{ is odd} \\ (n+2)/2 & \text{when } n \text{ is even} \end{cases} \tag{30}$$

A table of the coefficients a_1 to a_{22} to sixteen places (the last doubtful according to Jensen[10]) is given by Bourguet.[11] Schlömilch[12] also studied the properties of these constants and obtained a recursion formula and an infinite integral form. Nevertheless, the behavior of a_k in the limit of large values of k seems to be only poorly known; and correspondingly, an exact study of the properties of our series (29) is not possible on the basis of present knowledge. Applying the ratio test

[10] Ann. of Math., 17, 123 (1915). Translated by Gronwall.

[11] Acta Mathematica, 2, 261 (1883). Cf. also Bull. des Sci. Math., 16, 43 (1883), Davis (cf. 4e), v. I, p. 185.

[12] Z. für Math. und Physik, XXV, 103 (1880).

for absolute convergence, and using the familiar expression for the Bernoulli numbers[1,4b,4e,5a]

$$B_m = \frac{2}{(2\pi)^{2m}} \cdot \zeta(2m) \cdot (2m)! \tag{31}$$

where $\zeta(x)$ is the Riemann Zeta function, we obtain

$$\frac{1}{2\pi^2} \left| \operatorname*{Lim}_{m\to\infty} \frac{\zeta(2m+2)}{\zeta(2m)} \cdot m(2m+1) \cdot \frac{a_{2m-n+2}}{a_{2m-n}} \right| \tag{32}$$

Thus, in order to obtain absolute convergence

$$\operatorname*{Lim}_{m\to\infty} \left| \frac{a_{2m-n+2}}{a_{2m-n}} \right| \leqq \operatorname*{Lim}_{m\to\infty} \frac{2\pi^2}{m(2m+1)}$$

an extremely rigorous requirement, probably not fulfilled by a_k. We can however prove that the series is at least asymptotic, so that it is suitable for calculation, using the following theorem,[13] (using the following theorem,[13]) given by Knopp (Equation 300): "If $f(x)$ is of constant sign for $x > 0$, and, together with all its derivatives, tends monotonely to zero as $x \to \infty$, Euler's summation formula may be stated in the simplified form

$$\begin{aligned} \sum_{r=0}^{n} f(r) = \int_0^n f(x)\,dx - \tfrac{1}{2}[f(n) + f(0)] + \frac{B_1}{2!}[f'(n) - f'(0)] \\ + \cdots - (-1)^m \frac{B_m}{(2m)!}[f^{(2m-1)}(n) - f^{(2m-1)}(0)] \\ - (-1)^{m+1}\theta \frac{B_{m+1}}{(2m+2)!}[f^{(2m+1)}(n) - f^{(2m+1)}(0)] \end{aligned} \tag{33}$$

where $0 < \theta < 1$." Since, by equation (26) the required condition is fulfilled, we may terminate series (29) after any number of terms, with the assurance that the error thus committed will be less in absolute magnitude than the next term of the series.

(B) We now consider the infinite integral which we have developed above

$$I_n = \int_0^\infty y(x)\,dx; \qquad y(x) = \frac{x^n}{\Gamma(x+1)} \tag{34}$$

[13] *Theory and Application of Infinite Series*, translated by Young. Blackie (London, 1928).

Laplace[14] demonstrated that a definite integral of this type, taken between the limits of x which cause $y(x)$ to vanish, can be developed in an asymptotic series (subject to restrictions that are rather readily fulfilled),

$$\int_{x_1}^{x_2} y\,dx = Y\sqrt{\pi}\left(U + \frac{1}{2}\cdot\frac{1}{2!}\frac{d^2U^3}{dx^2} + \frac{1\cdot 3}{2^2}\cdot\frac{1}{4!}\frac{d^4}{dx^4}U^5 + \cdots + \frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2m-1)}{2^m\cdot(2m)!}\frac{d^{2m}}{dx^{2m}}U^{2m+1} + \cdots\right) \tag{35}$$

where Y is the maximum value of y between the assigned limits and

$$U = \left[-\frac{1}{2y}\cdot\frac{d^2y}{dx^2}\right]_{y=Y}^{-\frac{1}{2}} \tag{36}$$

In the limit then, we may use only the first term of this series,

$$\int_{x_1}^{x_2} y\,dx \sim \frac{\sqrt{2\pi}\,Y^{3/2}}{\sqrt{-\dfrac{d^2Y}{dx^2}}} \tag{37}$$

where $\dfrac{d^2Y}{dx^2}$ is written for $\left(\dfrac{d^2y}{dx^2}\right)_{y=Y}$. We shall not attempt to justify this formula of Laplace's further at this time, but shall proceed to employ it for the evaluation of K_n.

In this problem, y goes through a maximum when $\dfrac{dy}{dx} = 0$; but

$$y' = \frac{x^n}{\Gamma(x+1)}\left[\frac{n}{x} - \Psi(x+1)\right]$$

where

$$\Psi(x) = \frac{d}{dx}\ln\Gamma(x)$$

so that y is a maximum when $x = \alpha_n$, where α_n is the solution of the transcendental equation

$$\alpha_n = \frac{n}{\Psi(\alpha_n + 1)} \tag{38}$$

[14] Oeuvres, Tome 7, p. 103.

This equation can quite readily be solved numerically by successive approximations using the Newton-Raphson method. Thus if x_k is the k'th approximation to the root of

$$x\Psi(x+1) = n,$$

$$x_{k+1} = x_k + \frac{n - x_k\Psi(x_k+1)}{x_k\Psi'(x_k+1) + \Psi(x_k+1)}$$

$$= x_k - \frac{x_k\Psi^2(x_k+1) - n\Psi(x_k+1)}{\Psi^2(x_k+1) + n\Psi'(x_k+1)}$$

These equations lead to very rapid and convenient convergence in solving for α_n with tables of Ψ and Ψ'. Then

$$Y = \frac{\alpha_n^n}{\Gamma(\alpha_n+1)}; \qquad -\frac{d^2y}{dx^2} = \left\{\frac{\alpha_n^n}{\Gamma(\alpha_n+1)}\right\}\left[\frac{n}{\alpha_n^2} + \Psi'(\alpha_n+1)\right]$$

Substituting these values in (37) and (23a)

$$\int_0^\infty \frac{x^n\,dx}{\Gamma(x+1)} \sim \sqrt{2\pi}\,\frac{\alpha_n^{n+1}}{\Gamma(\alpha_n+1)}\cdot\frac{1}{\sqrt{n+\alpha_n^2\Psi'(\alpha_n+1)}} \tag{39}$$

and

$$K_n \sim \frac{\sqrt{2\pi}}{e}\cdot\frac{\alpha_n^{n+1}}{\Gamma(\alpha_n+1)}\cdot\frac{1}{\sqrt{n+\alpha_n^2\Psi'(\alpha_n+1)}} + \frac{1}{e}\sum_{m=m_0}(-1)^m\cdot\frac{B_m}{2m}\cdot a_{2m-n} \tag{40}$$

Generally speaking, the first term of equation 40, omitting the summation term entirely, is quite a fair approximation for the higher values of K_n. Thus while at $n = 10$, the result is about eighteen percent too high, at $n = 16$ where

$$\alpha_{16} = 7.63200\ 18618$$

we calculate

$$K_{16} \cong 10{,}481{,}629{,}910$$

which is greater than the true value of K_{16} (cf. Table I) by only about 0.014%. The first term of the series, $m = m_0 = 9$, is -1.76279; we note that, in this case at least, the summation term amounts to only a negligible fraction of the net error; so that the higher terms of the Laplace expansion (35) are much more important than the higher terms of (40).

For very large values of n, leading to large values of α_n, we may apply further approximations to (40). Introducing the approximations, valid for large n,

$$(41)\qquad \begin{cases} \Gamma(\alpha_n + 1) \to \sqrt{2\pi\alpha_n}\left(\dfrac{\alpha_n}{e}\right)^{\alpha_n} \\ \Psi(\alpha_n + 1) \to \ln(\alpha_n + 1) \\ \Psi'(\alpha_n + 1) \to 0 \end{cases}$$

we obtain

$$(42)\qquad K_n \sim \frac{\alpha_n^{n-\alpha_n} e^{\alpha_n - 1}}{\sqrt{\ln \alpha_n}}$$

If the restrictions of (41) are valid, α_n is the solution of the transcendental equation,

$$(43)\qquad \alpha_n = \frac{n}{\ln(\alpha_n + 1)}$$

and the solution of this equation can be obtained by iteration in the curious convergent form

$$(44)\qquad \alpha_n = \cfrac{n}{\ln\left(1 + \cfrac{n}{\ln\left(1 + \cfrac{n}{\ln\left(1 + \cfrac{n}{\ln(1 + \cdots}\right.}\right.}\right.}$$

If n is taken quite large, this is equivalent to

$$\alpha_n = \frac{n}{\ln n - \ln(\ln \alpha_n)}$$

or as a very crude approximation, valid only in the limit of extremely large n,

$$(45)\qquad \alpha_n \sim \frac{n}{\ln n}$$

With this value of α_n, equation (42) becomes

$$(46)\qquad K_n \sim \left[\frac{n e^{1/\ln n}}{\ln n}\right]^n$$

By taking logarithms of this equation, we obtain

$$\ln K_n \sim n\left[\ln\left(\frac{n}{e \ln n}\right) + \left\{1 + \frac{1}{\ln n}\right\}\right] \tag{47}$$

which is nearly identical with the asymptotic expansion for $\ln K_n$ given by Knopp[15]

$$\ln K_n \sim n\left[\ln\left(\frac{n}{e \ln n}\right) + \epsilon_n\right]$$

where, the author says $\epsilon_n \to 0$ as $n \to \infty$. We observe that from our calculations above ϵ_n approaches one rather than zero, a difference which is negligible, since the two expressions for $\ln K_n$ rapidly approach each other for large values of n.

By a direct comparison of the higher terms of the infinite series for K_x, $x^m e^x$, and e^{e^x}, we can show that in the limit of large x

$$x^m e^x < K_x < e^{e^x} \tag{48a}$$

for all finite values of m. A much tighter inequality follows from (46):

$$K_n \sim e^{n \ln n + n/\ln n - n \ln(\ln n)}$$

so that

$$e^{(n \ln n)(1-\eta_n)} < K_n < e^{(n \ln n)(1+\xi_n)}$$

where

$$\eta_n = \frac{1}{\ln n}\left[\frac{1}{\ln n}\right], \qquad \xi_n = \frac{1}{\ln n}\left[\ln (\ln n)\right]$$

Defining $g(n)$ as the difference between η_n and ξ_n, so that

$$g(n) = \frac{1}{\ln n}\left[\frac{1}{\ln n} - \ln (\ln n)\right]$$

we note that

$$\lim_{n \to \infty} g(n) = -\lim_{n \to \infty} \frac{\ln (\ln n)}{\ln n} = 0$$

and hence

$$\lim_{n \to \infty} \eta_n/\xi_n = 1$$

[15] Cf. Knopp (footnote 13), example 236.

Then letting the limiting value of η_n and ξ_n as n approaches infinity be ζ_n, we have

$$e^{(n \ln n)(1-\zeta_n)} < K_n < e^{(n \ln n)(1+\zeta_n)}$$

or

$$n^{n(1-\zeta_n)} < K_n < n^{n(1+\zeta_n)} \tag{48b}$$

where

$$\zeta_n = O\left(\frac{1}{\ln n}\right)^2$$

In conclusion, the writer wishes to sincerely thank Professor Philip Franklin for his inspiration and many helpful suggestions in the preparation of this paper, and Julius E. Epstein for his aid with the lengthy numerical calculations.

Appendix I

Subsequent to the preparation of the larger part of this paper, several references with bearing on the problems herein considered came to the attention of the writer. These results are here presented in brief.

1. Schwatt,[16] in studying the properties of the differential operator $\left(x\frac{d}{dx}\right)^n$, finds

$$\begin{aligned} K_n &= \sum_{k=1}^{n} \frac{(-1)^k}{k!} \sum_{\alpha=1}^{k} (-1)^\alpha \binom{k}{\alpha} \alpha^n \\ &= \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \left[\binom{k}{1} 1^n - \binom{k}{2} 2^n + \binom{k}{3} 3^n - \cdots + (-1)^{k-1} \binom{k}{k} k^n \right] \\ &= \sum_{k=1}^{n} (-1)^{k+1} \left[\frac{1^n}{1!} \cdot \frac{1}{(k-1)!} - \frac{2^n}{2!} \cdot \frac{1}{(k-2)!} + \frac{3^n}{3!} \cdot \frac{1}{(k-3)!} - \cdots \right. \\ &\qquad \left. \cdots (-1)^{k-1} \frac{k^n}{k!} \right] \end{aligned} \tag{I-1}$$

2. Chiellini[17] studies $\sum_{n=0}^{\infty} \frac{n^r}{n!}$ for integral r, by expanding in the inverse factorial series

$$\frac{n^r}{n!} = \sum_{k=1}^{r} \frac{b_{rk}}{(n-k)!} \tag{I-2}$$

[16] *An Introduction to the Operations with Series.* U. of Penn. Press, Philadelphia (1924).

[17] Boll. Un. Mat. It., X, 134 (1931).

where

I-3 $$b_{rk} = \frac{(-1)^k}{k!} \sum_{\alpha=1}^{k} (-1)^\alpha \binom{k}{\alpha} \alpha^r$$

I-4 $$b_{rk} = k b_{r-1,k} + b_{r-1,k-1}$$

so that

$$b_{r1} = 1^{r-1}$$

$$b_{r2} + b_{r1} = 2^{r-1}$$

$$b_{r3} + b_{r2} + \frac{b_{r1}}{2!} = \frac{3^{r-1}}{2!}$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

I-5 $$b_{rr} + b_{r,r-1} + \frac{b_{r,r-2}}{2!} + \frac{b_{r,r-3}}{3!} + \cdots + \frac{b_{r1}}{(r-1)!} = \frac{r^{r-1}}{(r-1)!},$$

finds

I-6 $$\sum_{n=1}^{\infty} \frac{n^r}{n!} = e \sum_{k=1}^{r} b_{rk} = b_r e$$

So that his b_r is identical with our K_r. He includes a table of b_r for integral r's from 1 to 10, with the misprints

$$4{,}140 = b_8 \neq 4{,}138$$

$$21{,}147 = b_9 \neq 22{,}147$$

3. Broggi[18] also studies $\sum_{n=1}^{\infty} \frac{n^h}{n!}$, h being integral and positive, by using the classical Stirling expansion

I-7 $$\prod_{p=0}^{n} \frac{1}{(x+p)} = \sum_{s=0}^{\infty} (-1)^s \frac{C_n^s}{x^{n+s}}$$

where

$$C_n^s = \frac{1}{n!} \sum_{r=0}^{n-1} (-1)^r (n-r)^{n+s} \binom{n}{r},$$

[18] Ist. Lombardo Rend. LXVI (IIs), 196 (1933).

so that the C_n^s's are what Nielsen[19] calls the Stirling numbers of the second kind, and designates $\mathfrak{C}_{n+1}^s$. Then from the known properties of these numbers, he demonstrates

$$\text{I-8} \qquad K_h = 1 + \frac{1\cdot(h-1)^h}{(h-1)!} + \frac{\left(1-\frac{1}{1!}\right)(h-2)^h}{(h-2)!} + \cdots + \cdots + \frac{1 - 1/1! + 1/2! - \cdots + (-1)^h 1/(h-2)!}{1!}$$

$$\text{I-9} \qquad K_h = C_h^0 + C_h^1 + C_{h-2}^2 + \cdots + C_1^{h-1}$$

and he proves that C_r^{r-k} is identical with Chiellini's b_{rk}. Finally he derives the asymptotic series

$$\text{I-10} \qquad \frac{1}{e}\int_0^1 t^{x-1}e^t\,dt = \frac{1}{x} + \sum_{n=1}^{\infty}(-1)^n\frac{K_n}{x^{n+1}}$$

4. Whitworth[20] shows that "the total number of ways in which n different things can be distributed into 1, 2, 3, 4, $\cdots$ or n indifferent parcels is $n!$ times the coefficient of x^n in the expansion of e^{e^x}/e," i.e. K_n. He further proves

$$\text{I-11} \qquad K_n = n!\sum_{t=1}^{n} N_t$$

where N_t is the number of t-partitions of n different things, and shows that N_t is $n!$ times the coefficient of x^n in the expansion of

$$\frac{(e^x-1)^t}{t!}$$

Glover[21] used equation I-11 above to compute the first six values of $K_n/n!$

[19] *Handbuch der Theorie der Gamma funktion.* Leipzig, 1906. 26, 27, 109.

[20] *Choice and Chance*, p. 96. Stechert, New York, 1925.

[21] *Tables of Applied Mathematics*, Ann Arbor, 1923. The writer is indebted to Dr. Glover for the reference to Whitworth above, and for his method of computing $K_n/n!$ (private communication).

5. Anderegg[22] proves

$$\text{I-12} \qquad K_n = \begin{vmatrix} 1 & -\binom{n-1}{0} & -\binom{n-1}{1} & \cdots & \binom{n-1}{n-2} \\ 1 & 1 & -\binom{n-2}{0} & \cdots & \binom{n-2}{n-3} \\ 1 & 0 & 1 & \cdots & \binom{n-3}{n-4} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & 1 & \binom{1}{0} \\ 1 & 0 & 0 & \cdots\ 0 & 1 \end{vmatrix}$$

and also gives the values of K_1 to K_{10}.

6. P. Epstein,[23] Krug, Tauchard, and Aitken have studied the arithmetical properties of K_n. Tauchard gives K_1 to K_{11}.

7. Bell[24] has contributed several papers on the numbers we have considered here. He shows[24a] that

$$\text{I-13} \qquad \xi_n^{(1)}(\text{Bell}) = K_n = \sum_{s=1}^{n} \frac{1}{(s-1)!}\left[\sum_{r=0}^{s-1} (-1)^r \binom{s-1}{r} (s-r)^{n-1}\right] \qquad (n > 0)$$

and discusses some of the arithmetic properties of these numbers. Further applications to the theory of numbers are also treated in the second[24b] and third[24c] papers. In the second paper he demonstrates

$$\text{I-14} \qquad K_n = \sum_{p=1}^{n} \frac{\Delta^p O^n}{p!}$$

where $\Delta^p O^n$ are the "differences of zero" discussed by various writers.[25] Equation I-14 is used, with tables of differences of zero, to find K_1

[22] Am. Math. Monthly, 8, 54 (1901) and 9, 11 (1901).

[23] a) P. Epstein: Archiv. der Math. und Physik., 8, 329 (1904–5).
b) Krug: Ibid., 9, 189 (1905).
c) Aitken: Math. Notes (Edinburgh), 28, 18 (1933).
d) Tauchard: Ann. Soc. Sci. Bruxelles A, 53, 21 (1933).

[24] a) Am. Math. Monthly, 41, 411 (1934).
b) Ann. of Math., 35, 264 (1934).
c) Ibid., 39, 539 (1938).

[25] Cf. Steffensen *Interpolation*, Williams & Wilkins, Baltimore (1927). See also Davis, v. II (footnote 4e), p. 210.

to K_{20}. The last three values in our Table I are taken from this paper —the rest were computed independently as indicated. The latest paper contains some interesting generalizations of the K-numbers, and discusses applications to computation; also an interesting interpretation of the significance of K_n in combinatory analysis: K_n "is the total number of possible rhyme schemes for a stanza of n verses."

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

THE MUTUAL INTERACTION OF PLASMA ELECTRONS

By W. R. Haseltine*

Abstract

This paper considers some of the processes taking place in a simplified model of the plasma or positive column region of an electrical discharge through gas at low pressure. Particular attention is given the direct mutual interaction of the electrons, and the effect of this on the distribution of velocities.

The first part deals with the question in a general way. To determine the velocity distribution function, two balance equations are set up. One expresses the conservation of energy, and the second the conservation of momentum parallel to the field. From these is derived one non-linear integral equation.

The geometry of the mutual scattering process is examined, and the necessary integrations are performed insofar as these depend only on this geometry. As an example the case of elastic sphere collisions is carried through to a point where integrations over the distribution function, f, become necessary.

In the second part the cross section function σ is derived for electron-electron scattering. As an interaction potential, the shielded Coulomb value $V = \frac{\epsilon^2}{r} e^{-r/a}$ was used.

In the third part an attempt is made to find functions which are solutions of the problem. Under the assumptions used, it is found that the function

$$f_0 = \frac{A}{2\pi w_0^3} e^{-g\frac{w^2}{w_0^2}-(1-g)\frac{w^4}{w_0^4}}$$

is a fairly good approximate solution.

It is found that experimental conditions can exist for which the theory predicts a value of the parameter g nearly unity. This corresponds to a nearly maxwellian distribution.

* This paper was submitted in partial fulfillment of the requirements for the degree of doctor of philosophy from the Massachusetts Institute of Technology. The author is at present a research fellow at the University of California.

I. **Introduction.** Many measurements made with probes in gas discharges have shown the existence of electronic velocity distributions of Maxwell-Boltzmann form, with very high average energies. Such results are, in particular, found when the electron density is high, and the net space charge (electrons plus positive ions) is low, as in the plasma of a low pressure arc. This form of distribution requires a very high rate of exchange of energy among the electrons.

In theoretical studies, this electronic interaction has been considered only by Langmuir (14) and Gabor (9). Of the two, Gabor's account is the more complete. It is now, however, satisfactory. He takes into account only the electrostatic phenomena, and is thus unable to compare their relative importance with that of collision processes and the macroscopic field. In his calculations, Gabor considers the electrons to be scattered by the fields of partially shielded, stationary positive ions. Because of the great mass of the positive ions, this cannot lead to as rapid a change in the velocity distribution as will the scattering of electrons by electrons. Furthermore, his methods of averaging are somewhat obscure.

At low electron densities, the electronic interactions can be neglected, as is shown by the quite good agreement of theory with experimental data (7, 21). Since these interactions increase with the square of the density, while the number of collisions with atoms increases with the first power, at some density the two effects will be equally important.

The object of this work is first, to find a means of calculating the mutual interactions of the electrons; second, to discover the exact conditions under which they become important; third, to find approximately the form of the velocity distribution in the cases in which neither process can be neglected.

The discussion will be limited to discharges between large, plane-parallel electrodes, so that only one direction in space can be singled out. It will be assumed that the positive ion and electron densities are equal, and that the drift current is small compared to the random electron current. It will be assumed further that the discharge is homogeneous in space, so that diffusion can be neglected.

The simplest mode of exchange of energy among the electrons is that between individual pairs, or by "collisions." Among more complicated processes is that suggested by Langmuir of "plasma oscillations." It is the mutual collisions which will be investigated in this paper.

For the sake of simplicity, attention will be confined to the following processes:

a, acceleration by the field;

b, elastic collisions with atoms;

c, mutual collision.

Collisions between electrons will be infrequent compared to collisions with atoms, but they are important. When two electrons collide, each is likely to suffer a considerable change in energy, while when one collides with an atom elastically, it loses on the average, only a fraction of the order $2m/M$ of its energy. Collisions with ions will be as infrequent as those with other electrons, and will result in as little an energy change as a collision with an atom. Accordingly, electron-ion collisions will be neglected.

Let $\mathfrak{N}$ be the electron density, and $\mathfrak{N}f d\gamma$ the number of electrons per cm^3 in an element $d\gamma$ of velocity space. f will depend on only the absolute velocity u, and the angle ω that this velocity makes with the field.

$$f = f(u, \cos\omega).$$

As f is nearly spherically symmetrical, it can be developed in terms of spherical harmonics, the first two terms giving:

$$f = f_0(u^2) + f_1(u^2)\cos\omega$$

and f_1 will be much smaller than f_0.

Following Boltzmann and Lorentz, the usual procedure to determine the distribution function would be to find the number of electrons entering and leaving each small element of velocity space per second on account of each type of process, and to set the sum equal to zero. In our problem, this involves too many successive integrations of the unknown function f, and is unmanageable.

Two equations are necessary to determine f_0 and f_1. These will be an energy balance and a momentum balance equation. The first can be obtained by requiring the net number of electrons leaving each central sphere of velocity space to be zero. The second is found by placing a similar requirement on the number leaving a region on one side of each plane perpendicular to the direction of the field. The first will mean that there will be just as many electrons having an initial velocity u, less than any given value w, which acquire in unit time a velocity greater than w, as there are having initial velocity greater than w and final velocity less than w. The second equation

expresses a similar balance of the changes in velocity component u_z parallel to the field.

II. **The General Equations.** Consider the energy balance equation, involving processes *a*, *b*, and *c*.

a. Electric field E. In time dt, an electron moving in a field E increases its velocity by an amount $\frac{\epsilon E}{m} dt \cos \omega$. The number per sec. that leave the central sphere of radius w will be:

$$\frac{2\pi\epsilon\mathfrak{N}E}{m} \int_0^\pi w^2(f_0 + f_1 \cos \omega) \cos \omega \sin \omega \, d\omega$$

or

$$\frac{4\pi\epsilon E\mathfrak{N}}{3m} w^2 f_1(w^2)$$

b. Elastic Collisions. An electron of velocity u, colliding elastically with an atom at rest, loses speed in amount $\delta u = \frac{m}{M} u(1 - \cos \theta)$, where θ is the angle between the initial and final velocity directions.

Let:

N = atomic density

ω = angle u makes with field E

ψ = corresponding azimuthal angle

$\sigma_a(u, \theta)$ = atomic cross section for elastic scattering at angle θ.

Then the number per sec. of electrons originally having a vectorial velocity in the element $u^2 \sin \omega d\omega d\psi$ which are scattered by angles between θ and $\theta + d\theta$ will be:

$$N\mathfrak{N}u^3 f \sin \omega \, d\omega \, d\psi \cdot 2\pi\sigma_a(u, \theta) \sin \theta \, d\theta.$$

The only electrons able to cross the sphere w due to elastic collisions will be those for which $w < u < w + \delta u$. The number per sec. entering such a sphere will be:

$$2\pi N\mathfrak{N} \int_0^\pi w^3 f \sin \omega \, d\omega \cdot 2\pi \int_0^\pi \delta u \sigma_a \sin \theta \, d\theta$$

If

$$Q = 2\pi \int_0^\pi \sigma_a(1 - \cos \theta) \sin \theta \, d\theta$$

this is

$$\frac{2\pi N\mathfrak{N}m}{M} w^4 Q \int_0^\pi (f_0 + f_1 \cos\omega) \sin\omega \, d\omega$$

Or, finally.

$$\frac{4\pi N\mathfrak{N}m}{M} w^4 Q f_0$$

c. Mutual Collisions. The mutual collisions of electrons are much more complicated.

Let:

u = initial velocity of electron considered,
v = " " " scattering electron,
U = velocity of center of gravity,
$\bar{u}$ = velocity of first electron relative to C. G.
u', $\bar{u}'$, v' = corresponding quantities after collision,
θ = scattering angle referred to C. G.
φ = corresponding azimuthal angle,
$\sigma(\bar{u}, \theta)$ = mutual scattering across section for angle θ,
χ = angle between directions of u and v,
η = corresponding azimuthal angle,
δ = angle between directions of U and $\bar{u}$,
λ = " " " " U and $\bar{u}'$,
μ = corresponding azimuthal angle,
ν = angle v makes with the field E.

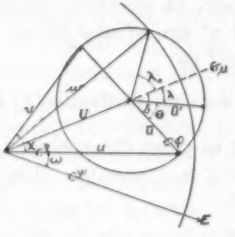


FIG. 1

Fig. 1 shows the geometry of a collision of an electron of velocity u with one of velocity v. Their relative velocity is $2\bar{u}$. Vectorially:

$$U = \tfrac{1}{2}(u + v), \qquad \bar{u} = \tfrac{1}{2}(u - v), \qquad u = \bar{u} + U.$$

Algebraically:

1 $$4\bar{u}^2 = u^2 + v^2 - 2uv\cos\chi$$

2 $$4U^2 = u^2 + v^2 + 2uv\cos\chi.$$

After collision the velocity of the first particle is: $u' = U + \bar{u}'$. $\bar{u} = \bar{u}'$, or u, u', v, v' all lie on the same sphere of radius $\bar{u}$. θ and φ are the polar angles of $\bar{u}'$ about $\bar{u}$ as an axis, and λ and μ are the polar angles of the same velocity about the axis U. The polar angles of v about the axis u are χ and η. σ depends on the *relative* velocity.

If u lies in the element of velocity space $d\gamma_u$, and v in the element $d\gamma_v$, the number of such collisions per sec. is:

$$2\bar{u}\sigma(\bar{u}, \theta)\sin\theta\, d\theta\, d\varphi\, \mathfrak{N}^2 f(u, \cos\omega) f(v, \cos\nu)\, d\gamma_u\, d\gamma_v.$$

To find the rate at which electrons of velocity $u < w$ are given a velocity $u' > w$ by mutual collision, this expression must be integrated: first, over all values of θ and φ on that portion of the sphere of radius $\bar{u}$ which lies outside the sphere of radius w; second, over a domain of γ_v (velocity space of second electron) such that the sphere $\bar{u}$ cuts the sphere w; third, over a domain of γ_u with $u < w$. The rate at which the reverse process takes place is found by integrating over θ and φ on that part of the sphere $\bar{u}$ *inside* the sphere w; over a part of the γ_v; and over a part of γ_u with $u > w$.

Because of symmetry, it is more convenient to use λ and μ than θ and φ.

3 $$\cos\theta = \cos\delta\cos\lambda - \sin\delta\sin\lambda\cos\mu.$$

For the purposes of integration, $\sin\theta\, d\theta\, d\varphi$ can be replaced by $\sin\lambda\, d\lambda\, d\mu$. Then μ will run from $-\pi$ to π, and λ from 0 to λ_0, where λ_0 is the value of λ on the circle of intersection of the spheres $\bar{u}$ and w. On this circle, vectorially: $\bar{u} + U = w$, or $\bar{u}^2 + U^2 + 2\bar{u}U\cos\lambda_0 = w^2$

4 $$\cos\lambda_0 = \frac{2w^2 - u^2 - v^2}{4U\bar{u}}$$

Also,

$$\bar{u}^2 + U^2 + 2\bar{u}U\cos\delta = u^2$$

whence,

5 $$\cos\delta = \frac{u^2 - v^2}{4U\bar{u}}$$

In the integration over γ_v, v, χ, and η can be used as variables, if $\cos \nu$ in $f(v, \cos \nu)$ is replaced by $\cos \omega \cos \chi + \sin \omega \sin \chi \cos \eta$.

Since $-1 \leq \cos \lambda_0 \leq 1$, and since, through U and $\bar{u}$, $\cos \lambda_0$ depends on χ, χ must be restricted. Expanding expression 4 for $\cos \lambda_0$,

$$6 \qquad \cos^2 \chi \leqq \frac{w^2(u^2 + v^2 - w^2)}{u^2 v^2}$$

Because $u^2 + v^2 \geq w^2$, $\cos^2 \chi \geq 0$. When u and v are on the same side of sphere w, (that is, when u and v are both greater or both less than w) the fraction is less than one, and it follows that $\chi_0 < \chi < \pi - \chi_0$. When u and v are on opposite sides of sphere w, there is no restriction on χ, as the fraction is greater than one. In this case it is possible to let $\chi_0 = 0$. Then the limits are always χ_0 and $\pi - \chi_0$.

The number of electrons leaving the sphere because of mutual collisions is:

$$I_1 = \mathfrak{N}^2 \int u^2 du \int_0^{\pi} f(u, \cos \omega) \sin \omega \, d\omega \int_0^{2\pi} d\psi \int v^2 dv \int_{\chi_0}^{\pi - \chi_0} \sin \chi \, d\chi$$

$$\cdot \int_{-\pi}^{\pi} f(u, \cos \nu) \, d\eta \cdot \int_0^{\lambda_0} 2\bar{u} \sin \lambda \, d\lambda \int_{-\pi}^{\pi} \sigma(\bar{u}, \theta) \, d\mu$$

Let:

$$7 \qquad B_1 = \int_0^{\lambda_0} \sin \lambda \, d\lambda \int_{-\pi}^{\pi} \sigma[\bar{u}, \theta(\delta, \lambda, \mu)] \, d\mu$$

The order of integration can be changed:

$$I_1 = 2\mathfrak{N}^2 \int u^2 du \int v^2 dv \int_{\chi_0}^{\pi - \chi_0} \bar{u}_1 B_1 \sin \chi \, d\chi \int_0^{\pi} f(u, \cos \omega) \sin \omega \, d\omega$$

$$\cdot \int_0^{2\pi} f(v, \cos \nu) \, d\eta \int_0^{2\pi} d\psi$$

Expanding f and substituting for $\cos \nu$:

$$\int_0^{2\pi} f(v, \cos \nu) \, d\eta = \int_0^{2\pi} [f_0(v^2) + f_1(v^2)(\cos \chi \cos \omega + \sin \chi \sin \omega \cos \eta)] \, d\eta$$

$$= 2\pi[f_0(v^2 + f_1(v^2) \cos \chi \cos \omega]$$

Taking the product of this result with $f(u, \cos\omega)$,

$$\int_0^\pi [f_0(u^2) + f_1(u^2)\cos\omega]\cdot[f_0(v^2) + f_1(v^2)\cos\chi\cos\omega]\sin\omega\, d\omega$$

$$= 2f_0(u^2)f_0(v^2) + \tfrac{2}{3}f_1(u^2)f_1(v^2)$$

f_1 is much smaller than f_0, so the product $f_1 f_1$ is negligible. Then:

$$8 \qquad I_1 = 16\pi^2\mathfrak{N}^2 \int f_0(u^2)u^2\, du \int f_0(v^2)v^2\, dv \int_{\chi_0}^{\pi-\chi_0} \bar{u}B_1 \sin\chi\, d\chi$$

Turning to the electrons entering the sphere due to mutual collisions, let:

$$9 \qquad B_2 = \int_{\lambda_0}^{\pi} \sin\lambda\, d\lambda \int_{-\pi}^{\pi} \sigma(\bar{u}, \theta)\, d\mu$$

The number entering is:

$$10 \qquad I_2 = 16\pi^2\mathfrak{N}^2 \int f_0(u^2)u^2\, du \int f_0(v^2)v^2\, dv \int_{\chi_0}^{\pi-\chi_0} \bar{u}B_2 \sin\chi\, d\chi$$

The three processes a, b, and c have now been calculated, and the energy equation can be written symbolically:

$$a - b + c = 0, \quad \text{or:}$$

$$11 \qquad \frac{4\pi\epsilon E\mathfrak{N}}{3} w^2 f_1 - \frac{4\pi m\mathfrak{N}NQ}{M} w^4 f_0 + (I_1 - I_2) = 0$$

This must be true for all values of w.

If the mutual scattering terms are omitted, one can derive from this equation, by one differentiation, Eq. 8 of Morse, Allis and Lamar (with diffusion left out).

$$\frac{\epsilon E}{2w}\frac{d}{dw}(w^2 f_1) = \frac{m}{M}\frac{3N}{2w}\frac{d}{dw}(w^4 Q f_0)$$

The mutual collisions of electrons will not be as important in the momentum balance equation as in the energy relation. In collisions with atoms electrons of energy H undergo a change of energy of the order $\frac{m}{M}H$. In corresponding encounters with other electrons, the change is of the order $H/2$. On the other hand, this is not true of the changes in momentum. Electrons are scattered by atoms nearly isotropically, and lose almost all of their directed momentum. The pair

interactions of electrons will be negligible for the purposes of this equation, because they are so much less frequent than collisions with atoms.

An equation corresponding to Eq. 11 can be set up for momentum balance. This leads to:

$$12 \qquad \frac{\epsilon E}{m}\frac{d}{dw}f_0 + NQwf_1 = 0$$

which is Eq. 7 of Morse, Allis, and Lamar.

So far, those integrations have been performed which involve only the geometry of the mutual scattering process. It is now necessary to consider the cross-section function σ. The true form of this is difficult to handle; so it is worthwhile to postpone its treatment, and to consider first the assumption that σ is constant. This would be true if the electrons behaved like hard spheres. The difficulties arising from the form of σ will be avoided, and it will be possible to show clearly the peculiarities of mutual collisions, as well as the method of handling them.

First, the energy and momentum equations can be combined to eliminate f_1.

$$13 \qquad \frac{-4\pi NQm\mathfrak{N}}{M}w^4 f_0 - \frac{4\pi\mathfrak{N}}{3NQ}\left(\frac{E\epsilon}{m}\right)^2 w\frac{d}{dw}f_0 + (I_1 - I_2) = 0$$

The four terms come from the elastic collisions, the field, and mutual collisions respectively.

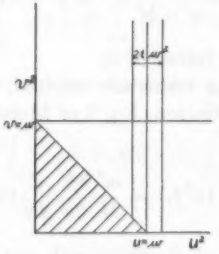


FIG. 2

III. **Constant Cross Section.** It is convenient to represent mutual collisions on a diagram of the v^2, u^2 plane. (Fig. 2) Each point stands for the velocities of a pair of particles. Because of the conservation of energy, every point P', representing the result of a collision between a pair P, will lie on a 45° diagonal line through P. Points in the shaded

region, for which $u^2 + v^2 < w^2$, cannot lead to collisions in which a particle crosses the sphere w. The lines $u = w$, and $v = w$ divide the plane into four quadrants, each of which must be treated separately. In $2(u < w, v > w)$ and $4(u > w$ and $v < w)$, the angle $\chi_0 = 0$, because u and v lie on opposite sides of sphere w. In $1(u < w, v < w)$ and $3(u > w, v > w)$, $0 < \chi_0 < \pi/2$.

Integration over quadrants 1 and 2 gives the electron flux across the line $u = w$ in an increasing direction; over 3 and 4, the flux in a decreasing direction.

Let:

$$B_1' = \int_{\chi_0}^{\pi-\chi_0} \bar{u} B_1 \sin\chi \, d\chi \qquad \text{Quad. 1}$$

$$B_2' = \int_0^{\pi} \bar{u} B_1 \sin\chi \, d\chi \qquad 2$$

14

$$B_3' = \int_{\chi_0}^{\pi-\chi_0} \bar{u} B_2 \sin\chi \, d\chi \qquad 3$$

$$B_4' = \int_0^{\pi} \bar{u} B_2 \sin\chi \, d\chi \qquad 4$$

From Eqs. 7 and 9,

$$B_1 = \int_0^{\lambda_0} \sin\lambda \, d\lambda \int_{-\pi}^{\pi} \sigma \, d\mu = 2\pi\sigma(1 - \cos\lambda_0)$$

$$B_2 = \int_{\lambda_0}^{\pi} \sin\lambda \, d\lambda \int_{-\pi}^{\pi} \sigma \, d\mu = 2\pi\sigma(1 + \cos\lambda_0)$$

Repeating Eqs. 1 and 2,

$$4\bar{u}^2 = u^2 + v^2 - 2uv\cos\chi$$

$$4U^2 = u^2 + v^2 + 2uv\cos\chi.$$

Eliminating $\cos\chi$

15 $$4U^2 = 2u^2 + 2v^2 - 4\bar{u}^2.$$

When u and v are held constant,

$$4\bar{u}\, d\bar{u} = uv \sin\chi \, d\chi$$

$$\sin\chi \, d\chi = \frac{4\bar{u}\, d\bar{u}}{uv}$$

Substituting for $\cos \lambda_0$ from Eq. 4,

$$B_1 = 2\pi\sigma\left[1 - \frac{2w^2 - u^2 - v^2}{uv}\right]$$

$$B_2 = 2\pi\sigma\left[1 + \frac{2w^2 - u^2 - v^2}{uv}\right]$$

Thus the integration over χ can be replaced by one over $\bar{u}$, if the proper limits are found. Let the limits on $\bar{u}$ be $\bar{u}_1$ and $\bar{u}_2$, $\bar{u}_1 < \bar{u}_2$, so that when,

$$\chi = \chi_0, \qquad \bar{u} = \bar{u}_1$$

$$\chi = \pi - \chi_0, \qquad \bar{u} = \bar{u}_2.$$

For Quads. 2 and 4, $\chi_0 = 0$, and thus:

$$4\bar{u}_1^2 = (u - v)^2, \qquad 4\bar{u}_2^2 = (u + v)^2.$$

Since $\bar{u}$ is essentially positive,

Quad.	2	4
$\bar{u}_1 =$	$(v - u)1/2,$	$(u - v)1/2$
$\bar{u}_2 =$	$(v + u)1/2,$	$(u + v)1/2.$

For quads. 1 and 3, from inequality 6,

$$\cos^2 \chi_0 = \frac{w^2(u^2 + v^2 - w^2)}{u^2v^2}$$

Using this value for $\cos \chi_0$,

$$4\bar{u}_1^2 = u^2 + v^2 + 2w\sqrt{u^2 + v^2 - w^2}$$

$$4\bar{u}_2^2 = u^2 + v^2 + 2w\sqrt{u^2 + v^2 - w^2}$$

Let $s^2 = u^2 + v^2 - w^2$. $1/2ms^2$ is the energy of the two particles in excess of that necessary to produce a collision in which one electron crosses the level w. In terms of s,

$$4\bar{u}_1^2 = (w - s)^2$$

$$4\bar{u}_2^2 = (w + s)^2.$$

This gives:

Quad.	1	3
$\bar{u}_1 =$	$(w - s)1/2$	$(s - w)1/2$
$\bar{u}_2 =$	$(w + s)1/2$	$(s + w)1/2.$

The integrals of Equs. 14 (over χ) can now be thrown into the forms:

$$B_1' = \frac{8\pi\sigma}{uv}\int_{(w-s)/2}^{(w+s)/2} \bar{u}\left[1 - \frac{w^2 - s^2}{2\bar{u}\sqrt{2u^2 + 2v^2 - 4\bar{u}^2}}\right]\bar{u}\,d\bar{u}$$

$$= \frac{8\pi\sigma}{uv}\left[\frac{\bar{u}^3}{3} + \frac{(w^2 - s^2)\sqrt{2u^2 + 2v^2 - 4\bar{u}^2}}{8}\right]_{(w-s)/2}^{(w+s)/2}$$

16a $$B_1' = \frac{8\pi\sigma}{3uv}\,s^3$$

Similarly:

$$B_3' = \frac{8\pi\sigma}{3uv}\int_{(s-w)/2}^{(s+w)/2} \bar{u}\left[1 + \frac{w^2 - s^2}{2\bar{u}\sqrt{2u^2 + 2v^2 - 4\bar{u}^2}}\right]\bar{u}\,d\bar{u}$$

16b $$B_3' = \frac{8\pi\sigma}{3uv}\,w^3$$

$$B_2' = \frac{8\pi\sigma}{uv}\int_{(v-u)/2}^{(v+u)/2} \bar{u}\left[1 - \frac{w^2 - s^2}{2\bar{u}\sqrt{2u^2 + 2v^2 - 4\bar{u}^2}}\right]\bar{u}\,d\bar{u}$$

16c′ $$B_2' = \frac{4\pi\sigma}{3uv}\,u(3v^2 - 3w^2 + 2u^2)$$

$$B_4' = \frac{8\pi\sigma}{3uv}\int_{(u-v)/2}^{(u+v)/2} \bar{u}\left[1 + \frac{w^2 - s^2}{2\bar{u}\sqrt{2u^2 + 2v^2 - 4\bar{u}^2}}\right]\bar{u}\,d\bar{u}$$

16d′ $$B_4' = \frac{4\pi\sigma}{3uv}\,v(3w^2 - v^2)$$

B_2' and B_4' can be simplified by taking advantage of the symmetry in u and v. They appear Eq. 13 as terms of $I_1 - I_2$, say $16\pi^2\mathfrak{N}^2 I'$, where

$$I' = \int_0^w f_0(u^2)u^2\,du\int_w^\infty f_0(v^2)v^2\,dv^2\,B_2'$$

$$- \int_w^\infty f_0(u^2)u^2\,du\int_0^w f_0(v^2)v^2\,dv^2\,B_4'$$

Substituting B_2' and B_4' from Eqs. 16′,

$$I' = \frac{4\pi\sigma}{3}\int_0^w f_0(u^2)u\,du\int_w^\infty f_0(v^2)v\,dvu(3v^2 - 3w^2 + 2u^2)$$

$$- \frac{4\pi\sigma}{3}\int_w^\infty f_0(u^2)u\,du\int_0^\infty f_0(v^2)v\,dvv(3w^2 - v^2)$$

Because of the symmetry of I' in u and v, $2u^2$ can be subtracted from the bracket in the first integrand, if $2v^2$ is subtracted from that in the second.

$$I' = 4\pi\sigma \int_0^w f_0(u^2)u^2\,du \int_w^\infty f_0(v^2)v^2\,dv\,\frac{(v^2 - w^2)}{v}$$

$$- 4\pi\sigma \int_w^\infty f_0(u^2)u^2\,du \int_0^w f_0(v^2)v^2\,dv\,\frac{(w^2 - v^2)}{u}$$

But this is equivalent to taking:

16c $$B_2' = \frac{4\pi\sigma}{v}(v^2 - w^2)$$

16d $$B_4' = \frac{4\pi\sigma}{u}(v^2 - w^2)$$

The functions B_i' derived here are much simpler than the ones to be calculated later for the shielded Coulomb field. They do show, however, the same symmetry properties in u and v.

If all processes except the mutual interactions are neglected, the solution of the balance equations will be a maxwellian distribution.

$$f_0 = Ae^{-u^2/w_0^2}.$$

This fact can be used to check the calculations. The u^2, v^2 plane can be divided into strips between the diagonal lines $u^2 + v^2 = c^2$ and $u^2 + v^2 = c^2 + \Delta c^2$. The first integration may be carried out along the length of these strips, instead of along lines parallel to the u^2 or v^2 axis. The maxwellian distribution function has the unique property of being constant along such a diagonal line. In other words, for this function and this path of integration, the f's disappear from the inner integral. It turns out that the B_i' functions can then be integrated. The result is zero for every value of c^2, showing that the mutual collisions do not change this distribution.

IV. **Calculation of Cross-Section.** The force between two electrons in the plasma is $\frac{l^2}{r^2}$ when they are very close, where r is their distance of separation. When r is larger, this force is reduced. Because the space is macroscopically neutral, at very large distances r the force between electrons will be zero. There is a space charge of positive ions, which, as far as the faster moving electrons are concerned, is

nearly uniformly distributed. Near each electron there will be, on average, a deficiency of other electrons, leaving an average net positive density which "shields" the electrons from each other.

A suitable approximation to the interaction potential is:

$$V = \frac{\epsilon^2}{r} e^{-r/\alpha}$$

This defines a potential "hill" whose radius is measured by α. The value to be assigned to α is uncertain. A lower limit is the mean distance of separation of the particles, $\mathfrak{N}^{-\frac{1}{3}}$. This may be of the order 10^{-4} cm. An obvious, but certainly excessive upper limit is the size of the apparatus, say 1 cm. It turns out that this enormous range for α is reflected in the results by an altogether smaller effect. α cannot, however, be infinite, for this would give the Coulomb potential, and with it, an infinite collision cross section. No value will be specified for this parameter at present.

The cross section $\sigma(\bar{u}, \theta)$ for this potential hill is found with the aid of quantum mechanics. The Born approximation can be used (29).

Let the wave function Ψ of the two electrons be a product of two functions, one representing the motion of the center of gravity, and the other, ψ, representing relative motion. If ψ be approximated by the sum of a plane wave, ψ_0, and a spherically scattered wave, ψ_1, then

$$\sigma = r^2 \frac{|\psi_1|^2}{|\psi_0|^2}$$

Let: $\psi_0 = e^{ikx_1 - ikx_2}$, where x_1 and x_2 are the distances of the two electrons from the center of gravity, and

$$k = \frac{2\pi m\bar{u}}{h}$$

Then the first approximation to the solution of Schrodinger's equation, with $V = \frac{\epsilon^2}{r} e^{-r/\alpha}$, gives:

$$\psi_1 = \frac{4\pi^2 me^2}{h^2} \cdot \frac{1}{4k^2 \sin^2 \frac{\theta}{2} + \frac{1}{\alpha^2}}$$

Or since $|\psi_0|^2 = 1$

$$\sigma = \left(\frac{4\pi^2 m\epsilon^2}{h^2}\right)^2 \frac{1}{\left(4k^2 \sin^2 \frac{\theta}{2} + \frac{1}{\alpha^2}\right)^2}$$

This can be written:

$$17 \qquad \sigma(\bar{u}, \theta) = \frac{\epsilon^4}{16m^2\bar{u}^4} \cdot \frac{1}{\left(\sin^2\frac{\theta}{2} + \bar{\beta}^2\right)^2}$$

where,

$$18 \qquad \bar{\beta}^2 = \frac{h^2}{(4\pi\alpha)^2 m^2 \bar{u}^2} = \left(\frac{\bar{\lambda}}{2\pi\alpha}\right)^2$$

$1/\bar{\beta}$ measures the size of the hill α in terms of $\bar{\lambda}$, the wave-length corresponding to the relative velocity $2\bar{u}$.

$\bar{\beta}^2$ is a very small number. Measured in volts,

$$\bar{\beta}^2 = \frac{.95 \cdot 10^{-16}}{v_r \alpha^2}$$

If the electronic density is 10^{12} per cc., and α is set equal to $\mathfrak{N}^{-\frac{1}{3}}$

$$\bar{\beta}^2 \simeq \frac{10^{-8}}{v_r}$$

and if α^2 is larger, $\bar{\beta}^2$ is by so much the smaller.

It is this term $\bar{\beta}^2$ in the denominator of σ, that prevents the cross section from becoming infinite at small angles, as it would for a Coulomb field. It will be the controlling part for distant encounters. Since it is so small, it will scarcely affect σ for large angles. These correspond to close encounters, and scattering by a nearly pure Coulomb field.

V(a). **Integration over the Angles.** In this section, the functions B and B' are calculated. Equation 17 can also be written:

$$\sigma = \frac{\epsilon^4}{4m^2\bar{u}^4} \cdot \frac{1}{(1 + 2\bar{\beta}^2 - \cos\theta)^2}$$

By equation 3 for $\cos\theta$, this becomes

$$\sigma = \frac{\epsilon^4}{4m^2\bar{u}^4} \cdot \frac{1}{(1 + 2\bar{\beta}^2 - \cos\delta\cos\lambda + \sin\delta\sin\lambda\cos\mu)^2}$$

When this is integrated over μ and λ, we obtain the functions B_1 and B_2 which follow.

$$B_1 = \frac{2\pi\epsilon^4}{16m^2\bar{u}^4(\bar{\beta}^4+\bar{\beta}^2)} \cdot\left[1 - \frac{(2\bar{\beta}^2+1)\cos\lambda_0 - \cos\delta}{\{4\bar{\beta}^4+4\bar{\beta}^2-2(2\bar{\beta}^2+1)\cos\delta\cos\lambda_0+\cos^2\delta+\cos^2\lambda_0\}^{\frac{1}{2}}}\right]$$

$$B_2 = \frac{2\pi\epsilon^4}{16m^2\bar{u}^4(\bar{\beta}^4+\bar{\beta}^2)} \cdot\left[1 + \frac{(2\bar{\beta}^2+1)\cos\lambda_0 - \cos\delta}{\{4\bar{\beta}^4+4\bar{\beta}^2-2(2\bar{\beta}^2+1)\cos\delta\cos\lambda_0+\cos^2\delta+\cos^2\lambda_0\}^{\frac{1}{2}}}\right]$$

B_2' is zero at $u = 0$, and $B_1' = 0$ along the line $u^2 + v^2 = w^2$.

These functions B are to be integrated over χ, the results are the B_i' functions. Exactly as was done in the case of constant cross section, an integration over $\bar{u}$ is substituted for that over χ.

$$\sin\chi\, d\chi = \frac{4\bar{u}\, d\bar{u}}{uv}$$

A new symbol "β" is defined,

$$\beta = \frac{2n}{4\pi\alpha m w} = \frac{2\bar{u}\bar{\beta}}{w} = \frac{\lambda_w}{2\pi\alpha}$$

where λ_w is the wave length corresponding to the velocity w. Since $\bar{\beta}$ is extremely small, "β" also is very small.

It is found upon evaluating the integrals over $\bar{u}$, that:

19a $$B_1' = \frac{2\pi\epsilon^4}{m^2w^2\beta^2uv}\left[s - \frac{w^2-u^2+w^2\beta^2}{2w\beta}\tan^{-1}\frac{2w\beta s}{w^2-u^2+w^2\beta^2}\right]$$

19c $$B_2' = \frac{2\pi\epsilon^4}{m^2w^2\beta^2uv}\left[u - \frac{w^2-u^2-w^2\beta^2}{2w\beta}\tan^{-1}\frac{2w\beta u}{w^2-u^2+w^2\beta^2}\right]$$

19b $$B_3' = \frac{2\pi\epsilon^4}{m^2w^2\beta^2uv}\left[w - \frac{u^2-w^2+w^2\beta^2}{2w\beta}\tan^{-1}\frac{2w\beta w}{u^2-w^2+w^2\beta^2}\right]$$

19d $$B_4' = \frac{2\pi\epsilon^4}{m^2w^2\beta^2uv}\left[v - \frac{u^2-w^2-w^2\beta^2}{2w\beta}\tan^{-1}\frac{2w\beta v}{u^2-w^2+w^2\beta^2}\right]$$

These may be compared with eq's. 16, which list the corresponding quantities for σ = constant.

The functions B_i' are large near the line $u = w$, and decrease rapidly away from it. In quadrants 2 and 3, they are independent of v. In quadrants 3 and 4, they are asymptotic to zero as u becomes large, and B_4' is zero when $v = 0$.

These functions are plotted in Fig. 3. In this figure, the value 0.1 was used for "β," so that the peaks are not so sharp.

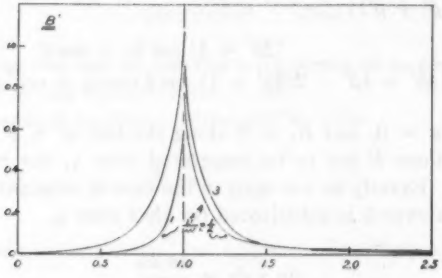


FIG. 3

V(b). **Integration over the Angles.** The B_i' functions would, when multiplied by the distribution function, be very difficult to integrate. Therefore approximations are made. Over most of the plane, it is satisfactory to expand the arc-tangent functions, retaining only terms through the order β^2. But this expansion breaks down for $|u^2 - w^2| < w^2\beta$. Hence, in a narrow strip about the $u = w$ line, some other method must be used. Over the width of this strip, the product $f(u^2)f(v^2)$ does not change anywhere nearly as rapidly as B_i'. So the distribution product is expanded, and the exact form of B_i' is retained.

When $w^2 - u^2 < w^2\beta$,

20a $$B_1' \simeq \frac{2\pi\epsilon^4}{m^2uv}\frac{4s^3}{3(w^2-u^2)^2}$$

20c $$B_2' \simeq \frac{2\pi\epsilon^4}{m^2uv}\left(\frac{4u^3}{3(w^2-u^2)^2} - \frac{2u}{w^2-u^2}\right)$$

20b $$B_3' \simeq \frac{2\pi\epsilon^4}{m^2uv}\frac{4w^3}{3(u^2-w^2)^2}$$

20d $$B_4' \simeq \frac{2\pi\epsilon^4}{m^2uv}\left(\frac{4v^3}{3(u^2-w^2)^2} - \frac{2v}{u^2-w^2}\right)$$

These expressions no longer contain "β," which is associated with small angle scattering. They represent that part of the flux across the level $u = w$, which is caused by collisions in which there is a comparatively large exchange of energy. They become infinite at $u = w$, while the more exact expressions do not. Outside of a strip $2tw^2$ (fig. 2) on each side of this line, they are very good approximations, ($t >> \beta$). Inside this strip the other approximation is used. The number "t" must be so chosen that on the one hand, $t >> \beta$, and on the other, so that the expansion of $f(u^2)f(v^2)$ is sufficiently precise inside the strip. It is shown later that this choice is not critical.

In the strip, instead of using the expansion,

$$f(u^2)f(v^2) \simeq f(w^2)f(v^2) + f'(w^2)f(v^2)(u^2 - w^2)$$

where

$$f'(w^2) = \frac{d}{dw^2} f(w^2)$$

it is advantageous to use one slightly different.

$$f(u^2)f(v^2) \simeq f(w^2)f(*v^2) + [f'(w^2)f(*v^2) - f(w^2)f(*v^2)](u^2 - w^2).$$

Here the point $(w^2, *v^2)$ is the intersection of the line $u^2 = w^2$ and a line of the family $u^2 + v^2 = \text{constant}$. These latter might be termed "lines of electron interaction," for any two points representing the velocities of a pair of electrons before and after collision respectively must lie on such a line. By using this expansion, a convenient cancellation of terms results.

Let S represent the integrals over the strip, and R those over the rest of the plane. Then

$$I_1 - I_2 = S + R. \tag{25}$$

To the order of approximation considered (terms in β^2):

$$\begin{aligned} R = \frac{8\pi^3 \epsilon^4 \mathfrak{N}^2}{m^2} \Bigg\{ & \int_0^{w^2} f_0(v^2)\, dv^2 \int_{w^2-v^2}^{w^2(1-t)} \tfrac{1}{2} f_0(u^2)\, \frac{S^3}{(w^2-u^2)^2}\, du^2 \\ & + \int_{w^2}^{\infty} f_0(v^2)\, dv^2 \int_0^{w^2(1-t)} f_0(u^2) \left[\frac{4u^3}{(w^2-u^2)^2} + \frac{2u}{w^2-u^2} \right] du^2 \\ & - \int_{w^2}^{\infty} f_0(v^2)\, dv^2 \int_{w^2(1+t)}^{\infty} f_0(u^2)\, \frac{4w^3}{3(u^2-w^2)^2}\, du^2 \\ & - \int_0^{w^2} f_0(v^2)\, dv^2 \int_{w^2(1+t)}^{\infty} f_0(v^2) \left[\frac{4v^2}{3(u^2-w^2)^2} + \frac{2v}{(u^2-w^2)} \right] dv^2 \Bigg\} \end{aligned}$$

and

$$S = \frac{8\pi^3 \epsilon^4 \mathfrak{N}^2}{m^2} \left\{ 2f_0(w^2) \left[w \int_{w^2}^{\infty} f_0(v^2)\, dv^2 - \int_0^{w^2} v f_0(v^2)\, dv^2 \right] \right.$$
$$+ \tfrac{8}{3} w^3 \left[\tfrac{1}{3} + \ln \frac{t}{2\beta} \right] \left[-f_0'(w^2) \int_{w^2}^{\infty} f_0(v^2)\, dv^2 + f(w^2) \int_{w^2}^{\infty} f_0'(v^2)\, dv^2 \right]$$
$$\left. + \tfrac{8}{3} \int_0^{w^2} \left[\tfrac{1}{3} + \ln \frac{tw}{2\beta v} \right] [-f_0'(w^2) f_0(v^2) + f_0(w^2) f_0'(v^2)] v^3\, dv^2 \right.$$

It is to be noticed that both t and β appear only in the logarithm.

It can easily be shown that both the exact values of β_i' listed in Eq's 19 and the approximate forms for S and R permit an equilibrium solution of maxwellian form.

Let

$$I_1 - I_2 = \frac{8\pi^3 \epsilon^4 \mathfrak{N}^2}{m^2 w^3} K$$

where K is a dimensionless factor, representing the integrals of eq. 25. The balance equation 13 is then:

$$\frac{-NQm}{M} w^4 f_0 - \frac{1}{3NQ} \left(\frac{E\epsilon}{m} \right)^2 w \frac{d}{dw} f_0 + \frac{2\pi^2 \epsilon^4 \mathfrak{N}}{m^2 w^3} K = 0$$

Letting

$$i = \frac{\pi \epsilon^4 \mathfrak{N}}{m^2 NQ}, \qquad \mathfrak{E} = \frac{E\epsilon}{mNQ}$$

this becomes

$$\frac{-m}{M} w^4 f_0 - \tfrac{1}{3} \mathfrak{E}^2 w \frac{d}{dw} f_0 + \frac{2\pi i}{w^3} K = 0 \qquad 26$$

$\mathfrak{E}$ corresponds to the parameter E/p of the Townsend theory. "i" is a measure of the intensity of ionization.

The function K can be separated into two parts: those terms containing the factor $\ln \frac{t}{2\beta}$; and the remainder. The former come from the narrow strip, and correspond roughly to small angle collisions, in which little energy is transferred. The latter come mostly from the regions of the u^2, v^2 plane outside of the strip of width tw^2, and represent close collisions with large energy exchanges.

K contains a number of rather complicated integrals, and a solution of equation 26 cannot be found directly. It is necessary to select a

function having the properties to be expected in a solution, and to test it by substituting it in the equation. For most choices, the integrals will have to be evaluated by numerical quadrature.

VI. **A Trial Function.** It is known that if the mutual interaction is neglected, the distribution function has the form:

$$f_0 = Ae^{-w^4/2w_0^4}$$

When only the pair collisions are considered, the distribution is maxwellian:

$$f_0 = Ae^{-w^2/w_0^2}$$

The function chosen as a trial solution is:

27
$$f_0 = \frac{A}{2\pi w_0^3} e^{-gw^2/w_0^2-(1-g)w^4/w_0^4}$$

$$= \frac{A}{2\pi w_0^3} e^{-\rho}$$

with

$$0 \leq g \leq 1.$$

At $g = 1$ or 0, this reduces to one or the other of those above. g and w_0 are parameters which are fo be determined so as to give the best approximation to a solution of eq. 26. g is a measure of the relative importance of electron-electron and electron-atom collisions. The number of electrons having speeds between w and $w + dw$ is: $4\pi\mathfrak{N}w^2 f_0(w^2)\, dw$. Taking the derivative,

$$\frac{d}{dw} w^2 f_0(w^2) = \frac{d}{dw} w^2 e^{-\rho}$$

$$= 2\left[w - g\frac{w^3}{w_0^2} - (1-g)\frac{w^5}{w_0^4}\right] e^{-\rho}$$

The derivative is zero at

$$1 - g\frac{w^2}{w_0^2} - (1-g)\frac{w^4}{w_0^4} = 0$$

or when $w^2 = w_0^2$. w_0 is therefore the most probable velocity. "A" is a normalization constant such that

$$1 = \frac{2A}{w_0^3}\int_0^\infty w^2 e^{-\rho}\, dw$$

letting

$$z = \frac{w^2}{w_0^2}, \text{ then: } 1 = A \int_0^\infty z^{\frac{1}{2}} e^{-\rho z - (1-g)z^2/2}\, dz$$

And A is a function of g alone.

Define:

$$\ln \frac{t}{2\beta} = \ln \frac{t}{2\beta_0} + \ln \frac{w}{w_0} \quad \text{or } \beta_0 = \frac{2h}{4\pi\alpha m w_0} = \frac{\lambda_0}{2\pi\alpha}$$

$$\beta_0 = \frac{2 - .975 \cdot 10^{-8}}{\alpha V_0^{\frac{1}{2}}}$$

V_0 is the energy $1/2 \cdot m w_0^2$ expressed in electron volts. If α is of the order 10^{-4}, and V_0 is a few volts, β_0 is of the order $1 \cdot 10^{-4}$. β_0 is introduced to separate the dependence of K on the parameters α, g, and w_0 from that on w.

Define:

$$K' = \frac{4\pi^2 w_0^5}{A^2 w^5} \frac{K}{e^{-\rho}}$$

K' is a dimensionless function of only $\ln \frac{t}{2a_0}$, g, and w/w_0.

Substituting in equation 26, and dividing by

$$\frac{w^2}{w_0^2} \frac{Ae^{-\rho}}{2\pi w_0^3}$$

$$-\frac{m w_0^4}{M} \frac{w^2}{w_0^2} + \frac{2}{3}\mathfrak{E}^2 \left[g + (1-g)\frac{w^2}{w_0^2} \right] + AiK' = 0$$

Rearranging:

$$28 \qquad iAK' = -\frac{2}{3}\mathfrak{E}^2 g + \frac{w^2}{w_0^2}\left[w_0^4 \frac{m}{M} - (1-g)\frac{2}{3}\mathfrak{E}^2 \right]$$

Because α and t only appear in the equation through their logarithms, the results are not critically dependent on the exact values given them. In the numerical calculations, t was chosen equal to 1/25. This is safely larger than β and is small enough so that the error in the expansion of $f(u^2)f(v^2)$ will be small.

Equation 28 expresses a relation that would be true if the trial

function chosen were a solution of the problem. It should be nearly true if this function is a good approximation. From the form of 28, it is seen that AK' should be nearly linear in w^2/w_0^2:

$$AK' = k_1 + k_2 w^2/w_0^2 .$$

In Fig. 4 values of K' as calculated numerically for $g = 0$ and $g = .732$ are plotted against w^2/w_0^2. These curves are seen to be nearly straight over the range containing most of the electrons ($w^2 < 4w_0^2$). The agreement is best for the larger values of g.

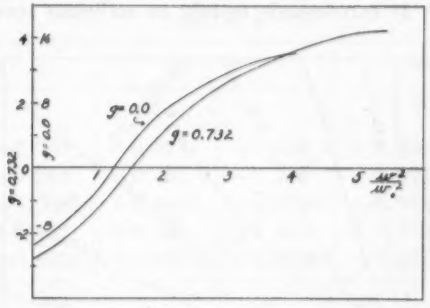


FIG. 4

From curves of this nature, one can find the values of g and w_0 which give functions most nearly correct for chosen values of i, $\mathfrak{E}$, and β_0. From eq. 28, follows eqs. 29 and 29′

$$29 \qquad ik_1 = -\frac{2}{3}\mathfrak{E}^2 g$$

$$29' \qquad ik_2 = -\frac{2}{3}\mathfrak{E}^2(1-g) + \frac{m}{M}w_0^4$$

Here k_1 and k_2 are functions of g, w_0 and $1n\beta_0$ being calculated by numerical quadrature.

These equations can be solved for g and w_0 in terms of the parameters i, $\mathfrak{E}$, and $\ln \beta_0$. In particular, g and $w_0^4/\mathfrak{E}^2$ are functions of $\log \beta_0$ and the ratio $\mathfrak{E}^2/i$ only. This is shown in Fig. 5, where lines of constant g and constant $w_0^4/\mathfrak{E}^2$ are plotted on the $\log \beta_0$, e^2/i plane.

Empirical equations for g and w_0 are:

$$g = \frac{0.26 + 2.46 \log \beta_0}{0.26 + 2.46 \log \beta_0 - \frac{2\mathfrak{E}^2}{3i}} \tag{30}$$

$$\frac{w_0^4}{\mathfrak{E}^2} = \frac{M}{m}\left[\frac{-2.92 - 2.39 \log \beta_0}{\mathfrak{E}^2/i} + \frac{2}{3}\right](1 - g) \tag{30'}$$

It has been mentioned that K' consists of two parts. One does not contain the parameter β_0. That is, it is independent of the size of the potential hill. It corresponds mainly to collisions involving a large

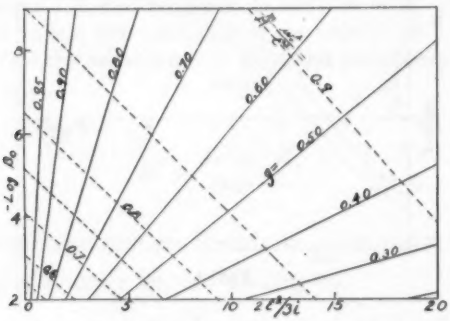


FIG. 5

change of energy. The second part, that contains the factor $\ln t/2\beta_0$, corresponds to glancing collisions with small energy exchange. These glancing collisions are much more numerous than the first kind, and it is interesting to see which type is the more important. Table I gives data for two values of β_0 that fall in the experimental range. Column 2 gives that part of the function $w^2/w_0^2K'e^{-\rho}$, which is independent of $\ln t/2\beta_0$. Columns 3 and 5 list that part of the same function containing the factor. Columns 4 and 6 give the sums of 2 with 3 and 5 respectively. It is seen that the larger part is that containing $1n\ \beta_0$, or that representing small angle collisions.

VII. It is interesting to see if conditions can exist for which this theory would predict a large value of g.

A search through the literature shows that there are very few papers

on the subject of gaseous conduction which present enough data to determine all three parameters. The required data are:

a, the gas used
b, the atomic density
c, the field strength
d, the electronic density
e, an estimate of the average energy.

Among the few papers giving all this data are, one by T. J. Killian (11), and one by A. H. van Gorcum (8). In Killian's work, the gas pressure was so low (a few bars) that the mean free path of the electrons

TABLE I

w^2/w_0^2		$\beta_0 = 10^{-4}$		$\beta_0 = 10^{-5}$	
1	2	3	4	5	6
0.342	−0.255	−0.373	−0.628	−0.530	−0.785
0.683	−0.281	−0.461	−0.742	−0.856	−0.937
1.366	−0.0091	−0.0695	−0.0786	−0.0989	−0.1080
2.732	+0.102	+0.222	+0.324	+0.316	+0.418
5.464	+0.00261	+0.00513	+0.00774	+0.00729	+0.00990

TABLE II

N	$\mathfrak{N}$	E	V_{av}
$1.67 \cdot 10^{17}$ cm^{-3}	$9.5 \cdot 10^{10}$ cm^{-3}	0.93 volts/cm	2.7 e.v.

was comparable to the tube diameter or larger. Plainly, this theory is inapplicable to such a case.

Van Gorcum reports work done on neon at a pressure of 4.7 mm. Hg. at 0°C. In the well developed plasma, he found by probe measurements that the electrons seemed to have very nearly a Maxwell distribution. The values found for the quantities listed above clustered about those in table II.

For neon, Q 0.28 cm^2. 10^{-15}.

If $\frac{eV}{300} = \frac{1}{2} m v_0^2$, then for a maxwellian distribution,

$$V_0 = \frac{2}{3} V_{av}$$

For this case, taking $\alpha = \mathfrak{N}^{-1} = 2.19 \cdot 10^{-4}$ we have;

$$\frac{2\mathfrak{C}^2}{3i} = 1.93, \qquad \beta_0 = 6.6 \cdot 10^{-5}, \qquad \log \beta_0 = -4.18$$

then from equation 30, $g = 0.84$.

This is quite a large value for g, and hence the mutual collisions are very important. If w_0 is calculated from eq. 30′, and $V_{0_{calc}} = \frac{1}{2} m v_{0_{calc}}^2$,

$$V_{0_{calc}} = 1.39 \text{ e.v.}$$

From the experiment,

$$V_0 = 1.8 \text{ e.v.}$$

These values are in fairly good agreement, considering the roughness of the approximations, and the fact that ionization has been neglected in the computations.

If α is chosen to equal 1 cm. it is found that:

$$g = 0.91, \quad \text{and} \quad V_{0_{calc}} = 1.45 \text{ e.v.}$$

The agreement between these two theoretical results is astonishing in view of the violently differing values of α that were used.

In discussing measurements made with probes in a plasma, it is usual to plot the logarithm of the electron current to the probe against the retarding voltage. For a maxwellian distribution, the resulting curve is a straight line:

$$\log i_p = c + V/V_0 .$$

In the general case, the current to the probe can be calculated from the equation:

$$i_p = \int_{\sqrt{2eV/m}}^{\infty} f(w^2) \left\{ w^2 - \frac{2eV}{m} \right\} w \, dw$$

It may be mentioned here that probe measurements are notoriously treacherous.

In Fig. 6, $\log i_p$ is plotted against V for the functions obtained by setting g equal to 0.5, 0.8, 1.0. The corresponding curves given by van Gorcum are quite a bit straighter than the one given here for $g = 0.8$. This would indicate that the mutual interactions are of even greater importance than is predicted by this theory. On the other hand, the

conditions in his tube were much more complicated than those assumed here. This fact may account for part of the divergence.

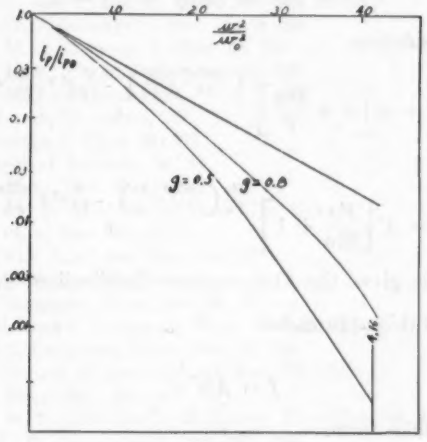


FIG. 6

VIII. Since the terms in $\ln \beta_0$ in the mutual interaction are so important, it is interesting to see what is obtained when they alone are retained. The balance equation then takes the form:

$$-\frac{m}{M} w^4 f - \frac{1}{3} \mathfrak{E}^2 w \frac{d}{dw} f$$

$$+ 2\pi i \frac{8}{3} \ln \frac{1}{2\beta_0} \left\{ \begin{array}{l} -w^3 f'(w^2) \int_{w^2}^{\infty} f(v^2)\, dv^2 + w^2 f(w^2) \int_{w^2}^{\infty} f'(v^2)\, dv^2 \\ -f'(w^2) \int_0^{w^2} v^3 f(v^2)\, dv^2 + f(w^2) \int_0^{w^2} v^3 f'(v^2)\, dv^2 \end{array} \right\} = 0$$

or, setting $x = w^2$

$$Cx^2 f(x) + Fx f'(x) + \mathfrak{M} \left\{ \begin{array}{l} x^3 f'(x) \int_x^{\infty} f(y)\, dy + f'(x) \int_0^x y^3 f(y)\, dy \\ + \frac{3}{4} f(x) \int_0^x y^3 f(y)\, dy \end{array} \right\} = 0$$

At large values of x, the first integral is very small, and the second and third are nearly constant. The equation is then nearly:

$$Cx^2f + Fxf' + \mathfrak{M}(af' + bf) = 0.$$

This has the solution

$$f = A\left[x + \frac{\mathfrak{M}a}{F}\right]^{-\left(\frac{C\mathfrak{M}^2a^2}{F^3} + \frac{\mathfrak{M}b}{Fa}\right)} e^{-\frac{C}{2F}x^2 + \frac{C\mathfrak{M}a}{F^2}x}$$

or

$$f = A'\left[\frac{Fx}{\mathfrak{M}a} + 1\right]^{-\frac{\mathfrak{M}a}{F^2}\left(\frac{C\mathfrak{M}ax}{F^2} + \frac{bx}{a}\right)} e^{-\frac{C}{2F}x^2 + \frac{C\mathfrak{M}a}{F^2}x}$$

If $\mathfrak{M} = 0$, this gives the Druyvesteyn distribution, and if $\mathfrak{M} \neq 0$, $F \to 0$, $\frac{C}{F} \to 0$ this approaches

$$f = A'e^{-\frac{b}{a}x}$$

or the maxwellian form.

IX. Conclusions. The scattering of electrons by the fields of other electrons in a plasma has been investigated, and a means devised for handling it. A trial distribution function inserted in the balance equation shows that this process has an important effect on the velocity distribution under certain conditions, namely; when E/NQ is small and $\mathfrak{N}/NQ$ is large. Such conditions actually occur experimentally, and for them this theory predicts that the distribution will be nearly maxwellian.

This trial function was not expected to be exact, but served to test the assumptions, and to indicate the relative importance of the various processes.

The scattering at small angles is more important than that at large angles. Or the velocity distribution is maintained principally by many small changes in energy rather than by less frequent large changes. The exact size α, chosen for the potential hill used in computing the scattering cross section is of relatively small importance.

The author wishes to express his gratitude for the unfailing encouragement, advice, and cooperation given him by Prof. William P. Allis.

Massachusetts Institute of Technology, Cambridge, Mass., and
University of California, Berkeley, California.

BIBLIOGRAPHY

1. Allen, H. W., Phys. Rev. **52**, 707
2. Allis, W. P., and H. W. Allen, Phys. Rev. **52**, 703
3. Compton, K. T., and C. Eckart, Phys. Rev. **25**, 139
4. Davydov, B., Phys. Zeitschr. d. Sov. Un. **9**, 433
5. Druyvestuyn, M. J., Zeitschr. f. Phys. **64**, 781
6. Druyvestuyn, M. J., Physikal. Zeitschr. **33**, 856
7. Druyvestuyn, M. J., Physica **3**, 65
8. Gorcum, H. H. van, Physica **3**, 207
9. Gabor, D., Zeitschr. f. Phys. **84**, 473
10. Gabor, D., Physikal. Zeitschr. **34**, 38
11. Killian, T. J., Phys. Rev. **35**, 1238
12. Langmuir, I., and H. Mott-Smith, Gen. Elec. Rev. **27**
13. Langmuir, I., Phys. Rev. **26**, 585
14. Langmuir, I., Nat. Acad. Sci. Proc. **14**, 627
15. Langmuir, I., Phys. Rev. **33**, 954
16. Found and I. Langmuir, Phys. Rev. **39**, 237
17. Mott-Smith, H., and I. Langmuir, Phys. Rev. **28**, 727
18. Tonks, L. and I. Langmuir, Phys. Rev. **33**, 195
19. McCurdy, W. H., and P. Dalton, Phys. Rev. **27**, 163
20. Morse, P. M., Phys. Rev. **31**, 1003
21. Morse, P. M., W. P. Allis, and E. S. Lamar, Phys. Rev. **48**, 412
22. Nottingham, W. B., Jour. Frank. Inst. **207**, 299
23. Pidduck, F. B., Proc. London Math. Soc. **15**, 89
24. Schottky, W., Physikal. Zeitschr. **25**, 342
25. Seeliger, R., Physikal. Zeitschr. **33**, 273
26. Sloane, R. H., and K. G. Emeléus, Phys. Rev. **44**, 333
27. Sibold, A., Helv. Phys. Acta **9**, 123
28. Smit, J. A., Physica **3**, 543
29. Sommerfeld, A., Atombau und Spektrallinien—Wellenmechanischer Ergan-zungsband
30. Thomson, J. J., The Conduction of Electricity Through Gases
31. Tonks, L. and W. P. Allis, Phys. Rev. **52**, 710
32. Tonks, L. and I. Langmuir, **34**, 896
33. Townsend, J. S., Phil. Mag. **11**, 1112
34. Uyterhoeven, W., Nat. Acad. Sci. Proc. **15**, 32

ALGEBRAIC EQUATIONS WITH COMPLEX COEFFICIENTS

By Frank L. Hitchcock

Equations with complex coefficients arise in electrical problems, for example in finding the transient currents generated when a polyphase induction motor is suddenly short-circuited through static condensers.

The following method is similar to that recently given for finding complex roots of equations with real coefficients.[1] For simplicity I shall first state the method for the cubic

$$x^3 + (a_2 + b_2 j)x^2 + (a_1 + b_1 j)x + a_0 + b_0 j = 0 \tag{1}$$

where the a's and b's are real and $j = \sqrt{-1}$. If this equation happens to have a *pure imaginary* root we can divide through by an imaginary factor $x - y_1 j$ where y_1 is real. If not, and if we call the real part of one of the roots r_1, we can suppose that by Horner's method all the roots have their real parts reduced by the amount r_1. We require a test to know when this has been accomplished.

Suppose the left member of the equation actually has a factor $x - yj$. Dividing through by this factor we find the remainder to be

$$a_0 - b_1 y - a_2 y^2 + (b_0 + a_1 y - b_2 y^2 - y^3)j. \tag{2}$$

If this remainder is zero we have simultaneously

$$y^3 + b_2 y^2 - a_1 y - b_0 = 0 \tag{3}$$

and

$$a_2 y^2 + b_1 y - a_0 = 0. \tag{4}$$

Hence the equations (3) and (4) must have a common solution $y = y_1$. This is the same as saying that the left members of these two equations have a common factor $y - y_1$, in which case the left side of (1) will have a factor $x - y_1 j$. We can find the common factor $y - y_1$, if it exists, by the usual process of finding the greatest common divisor of two polynomials.

If, as of course will usually happen, there is no common factor, the process of finding the greatest common divisor will yield a remainder

[1] This Journal, Vol. XVII, No. 2, June 1938, p. 55.

which is a real number. This remainder will, in general, have opposite signs according to whether, by Horner's method, we have reduced the real parts of the roots not quite enough or a little too much.

As an example take the cubic

$$x^3 + (4 - j)x^2 + (3 + 2j)x - 3 + 4j = 0. \tag{5}$$

Here we have

$$a_0 = -3 \quad a_1 = 3 \quad a_2 = 4$$

$$b_0 = 4 \quad b_1 = 2 \quad b_2 = -1$$

so that the test equations (3) and (4) become

$$4y^2 + 2y + 3 = 0, \qquad y^3 - y^2 - 3y - 4 = 0.$$

For convenience we divide the first of these equations through by $+4$ giving $y^2 + 0.5y + 0.75 = 0$, and then find the G.C.D. of the left sides.

$$\begin{array}{lll} y^2 + 0.5y + 0.75) & y^3 - y^2 - 3.\ y - 4. & (y - 1.5 \\ & \underline{y^3 + 0.5y^2 + 0.75y} & \\ & -1.5y^2 - 3.75y - 4. & \\ & \underline{-1.5y^2 - 0.75y - 1.125} & \\ & -3.00y - 2.875 = \text{first remainder.} & \end{array}$$

For convenience we divide this first remainder through by $+3$ giving $-y - 0.958$ for a new divisor. (We must not divide through by a *negative* number since this might alter the sign of the final remainder.) The work continues:

$$\begin{array}{lll} -y - 0.958) & y^2 + 0.5\ y + 0.75 & (-y + 0.458 \\ & \underline{y^2 + 0.958y} & \\ & -0.458y + 0.75 & \\ & \underline{-0.458y - .439} & \\ & +1.189 = \text{final remainder.} & \end{array}$$

We now merely note that the final remainder is positive. Returning to the original equation (5) we may try reducing the real parts of the roots by the amount $+1$. With complex coefficients there is no difference in principle when applying Horner's method; we merely keep a

kind of double entry in the computation, the columns of imaginary values being headed by 'j', if we wish.

	j		j		j	
4.	−1.	3.	2.	−3.	4.	(+1.
1.		5.	−1.	8.	1.	
5.	−1.	8.	1.	5.	5.	
1.		6.	−1.			
6.	−1.	14.	0.			
1.						
7.	−1.					

Thus the new equation, whose roots all have their real parts less by +1 than those of the original equation, is

$$x^3 + (7 - j)x^2 + 14x + (5 + 5j) = 0 \tag{6}$$

For this new equation we have

$$a_0 = 5 \quad a_1 = 14 \quad a_2 = 7$$

$$b_0 = 5 \quad b_1 = 0 \quad b_2 = -1$$

so that the new test equations (3) and (4) are

$$7y^2 - 5 = 0, \qquad y^3 - y^2 - 14y - 5 = 0.$$

We divide the quadratic through by +7, giving $y^2 - 0.714 = 0$, and then find the G.C.D. of the left sides.

$$\begin{array}{l|l} y^2 - 0.714) & y^3 - y^2 - 14. \quad y - 5. \qquad (y - 1 \\ & \underline{y^3 \qquad - 0.714y} \\ & \quad - y^2 - 13.286y - 5. \\ & \quad \underline{- y^2 \qquad + .714} \\ & \qquad -13.286y - 5.714 \end{array}$$

Divide this first remainder through by +13.286 giving $-y - 0.430$, which is to be divided into the previous divisor,

$$\begin{array}{l|l} -y - 0.430) & y^2 \qquad - 0.714 \quad (-y + 0.430 \\ & \underline{y^2 + 0.430y} \\ & \quad -0.430y - 0.714 \\ & \quad \underline{-0.430y - 0.185} \\ & \qquad - 0.529. \end{array}$$

The final remainder is now negative, whereas before it was positive. This proves that the original equation (5) has a root whose real part is between zero and +1.

Returning to the original equation it is natural next to try reduction of the real parts of the roots by the amount +0.5. The sequel will then show whether the real part of the root in question is greater or less than 0.5.

4.	−1.	3.	2.	−3.	4.	(0.5
.5		2.25	− .5	2.625	.75	
4.5	−1.	5.25	1.5	−0.375	4.75	
.5		2.50	− .5			
5.0	−1.	7.75	+1.0			
.5						
5.5	−1					

Thus the new equation, whose roots all have their real parts less by 0.5 than those of the original equation is

$$x^3 + (5.5 - j)x^2 + (7.75 + j)x - 0.375 + 4.75j = 0. \qquad (7)$$

Here we have

$$a_0 = -0.375 \quad a_1 = 7.75 \quad a_2 = 5.5$$

$$b_0 = 4.75 \quad b_1 = 1. \quad b_2 = -1.$$

so that the test equations (3) and (4) now become

$$5.5y^2 + y + 0.375 = 0, \qquad y^3 - y^2 - 7.75y - 4.75 = 0.$$

We divide the quadratic through by 5.5, giving $y^2 + 0.182y + 0.068$ and then find the G.C.D. of the left sides.

$$\begin{array}{l} y^2 + 0.182y + 0.068) \; y^3 - \quad y^2 - 7.75\,y - 4.75 \; (y - 1.182 \\ \qquad\qquad\qquad\quad y^3 + 0.182y^2 + \quad .068y \\ \qquad\qquad\qquad\quad \overline{- 1.182y^2 - 7.818y - 4.75} \\ \qquad\qquad\qquad\quad - 1.182y^2 - \quad .215y - \quad .080 \\ \qquad\qquad\qquad\quad \overline{\qquad\qquad - 7.603y - 4.67} \end{array}$$

Divide this first remainder through by $+7.603$ giving $-y - 0.614$, which is to be divided into the previous divisor,

$$\begin{array}{r} -y - 0.614) \; y^2 + 0.182y + 0.068 \; (-y + 0.432 \\ \underline{y^2 + .614y} \\ -\ .432y + .068 \\ \underline{-\ .432y - .265} \\ +\ .333 \end{array}$$

a positive final remainder. This proves that the original equation has a root whose real part is between $+0.5$ and $+1.0$.

We may now return to the original equation and reduce the roots by 0.7.

						(0.7
4.	−1.	3.	2.	−3.	4.	
.7		3.29	− .7	4.403	.91	
4.7	−1.	6.29	1.3	1.403	4.91	
.7		3.78	− .7			
5.4	−1.	10.07	.6			
.7						
6.1	−1					

Thus the equation whose roots all have their real parts less by 0.7 than those of the original equation (5) is

$$x^3 + (6.1 - j)x^2 + (10.07 + 0.6j)x + 1.403 + 4.91j = 0. \quad (8)$$

Here we have

$$a_0 = 1.403 \qquad a_1 = 10.07 \qquad a_2 = 6.1$$

$$b_0 = 4.91 \qquad b_1 = 0.6 \qquad b_2 = -1,$$

so that the test equations (3) and (4) now are

$$6.1y^2 + 0.6y - 1.403 = 0, \qquad y^3 - y^2 - 10.07y - 4.91 = 0.$$

We divide the quadratic through by 6.1 giving $y^2 + .0983606y - 0.23 = 0$, and then find the G.C.D. of the left sides,

$$\begin{array}{r} y^2 + .0984y - .23) \; y^3 - \quad y^2 - 10.07 \; y - 4.91 \quad (y - 1.0984 \\ \underline{y^3 + .0984y^2 - \quad .23 \; y} \\ -\ 1.0984y^2 - 9.84 \; y - 4.91 \\ \underline{-\ 1.0984y^2 - \quad .1081y + \quad .2526} \\ -\ 9.7319y - 5.1626 \end{array}$$

Divide this first remainder through by $+9.7319$ giving $-y - 0.5305$, which is to be divided into the previous divisor,

$$\begin{array}{rl} -y - 0.5305) & y^2 + .0984y - 0.23 \quad (-y + .4321 \\ & \underline{y^2 + .5305y} \\ & \quad - .4321y - 0.23 \\ & \quad \underline{- .4321y - \ .2292} \\ & \qquad - \ .0008 \end{array}$$

a small negative remainder. This proves that the original equation (5) has a root with real part a little less than 0.7, but probably only a little less, because the remainder is so small.

We could continue in this way, testing, perhaps, 0.69, which would be done most easily by taking (8) and increasing its roots by 0.01. It is more rapid, however, to procede as follows. If the remainder in the last division had been exactly zero this would have meant that the equation (8) had a factor precisely $y + 0.5305j$ and hence the original equation a root precisely $0.7 - 0.5305j$. We can conclude that this value is a fair approximation to a root.

Equation (8) has thus one root whose real part is nearly zero. Let any cubic known to have one root with real part near zero be

$$x^3 + (A_2 + B_2j)x^2 + (A_1 + B_1j)x + A_0 + B_0 = 0. \tag{9}$$

Suppose the imaginary part of the root in question is known to be approximately kj. If we move the origin parallel to the imaginary axis by an amount kj the resulting equation will have a root nearly zero. This is the same as subtracting the amount kj from all the roots by Horner's method. The first remainder is found to be $s + tj$ where

$$s = A_0 - B_1k - A_2k^2 \tag{10}$$

$$t = B_0 + A_1k - B_2k^2 - k^3 \tag{11}$$

which are, in form, like the real and imaginary parts of (2). The second remainder is $u + vj$ where

$$u = A_1 - 2B_2k - 3k^2 \tag{12}$$

$$v = B_1 + 2A_2k. \tag{13}$$

These formulas may be obtained in another way. Call the left side of (9) $f(x)$ and write $x' + kj$ for x. Expanding by Taylor's theorem we have

$$f(x' + kj) = f(kj) + x'f'(kj) = s + tj + x'(u + vj) = 0 \tag{14}$$

if we suppose the new origin is so near the actual root that x'^2 may be neglected. It is easy to check that u, v, s, and t obtained by this method agree with Horner's method. Solving for x' we have

$$x' - \frac{su + tv + (tu - sv)j}{u^2 + v^2} \tag{15}$$

which is the amount to be added to the approximate root to obtain a better root. This is "Newton's method".

Returning to (8) we have

$$A_0 = 1.403, \quad A_1 = 10.07, \quad A_2 = 6.1$$

$$B_0 = 4.91\ , \quad B_1 = 0.6\ , \quad B_2 = -1.$$

which are, of course, the same as a_0, b_0, etc. already set down for equation (8). (The only object in using A_0, B_0, etc. in (9) is to emphasize the important point that the use of s, t, u, and v to obtain x' has no meaning *except* when we start with an equation having the real part of one root very small.) Finding s, t, u, and v and then x', by aid of a computing machine, it was found that

$$x' = -0.0004\ 5098 - 0.0001\ 5214\ j. \tag{16}$$

This proves that the approximate root $0.7 - 0.5305j$ was correct to three decimal places. If greater accuracy is not wanted nothing more needs to be done on this root. If we want more places we may suppose that x' has significance to three more places, which is probable but not certain. Adding to the value $0.7 - 0.5305j$ the correction $-0.000451 - 0.000152j$ we have

$$+ 0.699549 - 0.530652j \tag{17}$$

as a better value of this root.

An easy check is to return to (8) and increase the real parts of the roots by the amount of the correction 0.000451 (using a machine). Then carrying out the G.C.D. process we should find, first, that the final remainder is very small, and, second, that the last divisor is $\pm (y + 0.530652)$ rather closely. (Actually we get $-y - 0.5306\ 5223$ and remainder $- .0000\ 0012$.)

If complete rigor in estimating the error is needed, we can attain this by examining the change of sign in the final remainder.

Considering now the general case of an equation of any degree n,

$$x^n + (a_{n-1} + b_{n-1}j)x^{n-1} + \cdots + a_0 + b_0 j = 0. \tag{18}$$

The remainder on dividing the left side by $x - yj$ is found by substituting yj in place of x and so is

$$(yj)^n + (a_{n-1} + b_{n-1}j)(yj)^{n-1} + \cdots + a_0 + b_0 j. \tag{19}$$

Hence the test equations are

$$a_0 - b_1 y - a_2 y^2 + b_3 y^3 + a_4 y^4 - b_5 y^5 - a_6 y^6 + \cdots = 0 \tag{20}$$

and

$$b_0 + a_1 y - b_2 y^2 - a_3 y^3 + b_4 y^4 + a_5 y^5 - b_6 y^6 - \cdots = 0. \tag{21}$$

If n is odd (20) will terminate with a term in y_{n-1} and (21) will terminate with a term in y^n and coefficient ± 1. If n is even (20) will terminate with a term in y^n and coefficient ± 1 while (21) will terminate with a term in y^{n-1}. In each case the coefficients are alternately an 'a' and a 'b' (with proper sign) except for the term in y^n. The signs begin as indicated and continue alternately two plus signs and then two minus signs.

The general formulas for s and v are of the same form as (20) and (21),

$$s = A_0 - B_1 k - A_2 k^2 + B_3 k^3 + A_4 k^4 - B_5 k^5 - A_6 k^6 + \cdots \tag{22}$$

$$t = B_0 + A_1 k - B_2 k^2 - A_3 k^3 + B_4 k^4 + A_5 k^5 - B_6 k^6 - \cdots \tag{23}$$

We find $u + vj$ by putting kj for x in $f'(x)$, giving

$$u = A_1 - 2B_2 k - 3A_3 k^2 + \cdots \tag{24}$$

$$v = B_1 + 2A_2 k - 3B_3 k^2 - \cdots \tag{25}$$

where, if n is odd, the series for u terminates in a term $+nk^{n-1}$ and v in a term in k^{n-2}; but if n is even the situation is reversed. The signs fall as in s and t.

In case the given equation has two or more roots having their real parts nearly equal, the foregoing use of the G.C.D. would not readily separate these roots. Actually this is the situation likely to arise in a practical case, where all the real parts may be quite small, the principle problem being to find their magnitude, particularly such real parts as are positive. The imaginary parts are likely to be far from equal. In such a situation the remedy is simple. If we put $-jx'$ in place of x in the given equation we induce a rotation of the complex plane so that we interchange real and imaginary parts of all roots. We now have two or more imaginary parts nearly equal, which does no harm, while the real parts of the new equation are far from equal. Then we

proceed by the G.C.D. method, or with s, t, u, v, just as in the example worked out above.

As an illustration, the following equation, submitted through the kindness of Professor W. V. Lyon, occurred in an actual problem,

$$x^3 + (0.4 - j)x^2 + (6.67 - 0.2j)x + 0.0667 - 6.67j = 0. \quad (26)$$

It is known from the nature of the problem, and can be easily verified, that the real parts of all the roots are numerically small. If we put $-jx'$ in place of x the equation becomes

$$x'^3 + (1 + 0.4j)x'^2 + (-6.67 + 0.2j)x' - 6.67 - 0.0667j = 0 \quad (27)$$

which may be solved by the method of the foregoing pages, for the small real parts have now been replaced by small imaginary parts, so that the new real parts may be readily separated by the G.C.D. method.

When the values of x' have been found the values of x, viz. the roots of the original equation, are of course found from $x = -jx'$. The results were found to be

$$-0.246 + 2.608j, \quad -0.177 - 2.606j, \quad \text{and} \quad +0.0233 + 0.9975j \quad (28)$$

In conclusion it may be remarked that the foregoing method still holds for equations with real coefficients, so far as the problem of finding complex roots is concerned. We merely have all the b's zero. In fact if we take an equation of even degree with real coefficients, the test equations of the present paper reduce to those of the former paper.

Massachusetts Institute of Technology.

Reprinted from JOURNAL OF MATHEMATICS AND PHYSICS
Vol. XVIII, No. 3, July, 1939

ON ASYMPTOTIC ORBITS IN THE THEORY OF PRIMARY COSMIC RADIATION

BY ALFREDO BAÑOS, JR.*

1. **Introduction.** The investigations reported here deal with certain theoretical aspects of the problem of the motion of charged particles in the field of a magnetic dipole which concern directly the Lemaître-Vallarta theory of the charged component of cosmic radiation. Briefly stated, then, the problem discussed here refers to the calculation, for a given value of Störmer's parameter γ_1, of a family of trajectories whose members are asymptotic to a given symmetric, unstable, periodic orbit.

The theory of the charged component of cosmic radiation is based on the hypothesis of the existence, at large distances from the earth, of a homogeneous and isotropic distribution of charged particles. Lemaître and Vallarta[1] have shown that, by virtue of the earth's magnetic field, cosmic rays of a given energy do not reach all points on the surface of the earth. In general, at a given magnetic latitude, charged particles of a given energy are capable of reaching the earth only along certain allowed directions, the integrated totality of which fills a cone of somewhat complex and irregular shape.[2,3] Furthermore, as a corollary to Liouville's theorem, it follows that the intensity of the charged component of cosmic radiation per unit solid angle within the cone of allowed directions, for a given energy, is the same as the intensity at large distances from the earth.[1] In the theory of cosmic radiation, therefore, the fundamental problem consists of the determination, for every latitude and for every energy, of the shape of the cone of allowed directions whose generators are formed by trajectories separating orbits coming from infinity from orbits which originate at some other point of the earth. An important class of such generators is constituted by

* Fellow of the John Simon Guggenheim Memorial Foundation, now Director of the Institute of Physics of the National University of Mexico, Mexico City.

[1] G. Lemaître and M. S. Vallarta, Phys. Rev., 43, 88 (1933); G. Lemaître, Ann. de la Soc. Sci. de Bruxelles, A54, 162 (1935).

[2] G. Lemaître and M. S. Vallarta, Phys. Rev., 49, 720 (1936).

[3] G. Lemaître and M. S. Vallarta, Phys. Rev., 50, 500 (1936).

orbits which are asymptotic to certain unstable periodic orbits originally discovered by Störmer.[4]

In a recent paper by Lemaître and Vallarta,[2] to which reference should be made for a complete statement of the problem treated and of their method of attack, results were presented from the analysis of some three hundred trajectories, asymptotic to a known family of unstable periodic orbits, as obtained by means of Bush's differential analyzer.[5] These results were submitted then with reservations pending a critical examination of the precision associated with machine trajectories. In a later article,[3] Lemaître and Vallarta presented a full discussion of their method for the determination of asymptotic trajectories by means of the differential analyzer together with a brief outline of the method of calculation of asymptotic trajectories which they developed, using a generalized method of numerical integration, to afford a comparison between machine trajectories and computed trajectories.

To ascertain the magnitude of possible systematic errors in the mechanical determination of asymptotic trajectories Lemaître and Vallarta computed, for $\gamma_1 = 0.93$, the complete inner family of asymptotic orbits, which they represented by means of two trigonometric series whose amplitudes, as functions of σ, they determined by numerical integration.[6] The purpose of the present investigations, then, was two-fold: to carry out, under more adverse conditions, the Lemaître-Vallarta method of numerical integration, bisecting the interval of integration as many times as might be required to exhaust fully the practical possibilities of the method; and, by carefully drawing the complete family of asymptotic orbits for $\gamma_1 = 0.85$, to afford a closer and more critical analysis of the precision associated with machine trajectories than had been possible heretofore.

2. **Formulation of the Problem.** The complete formulation of the problem involved in the present discussion must necessarily be pre-

[4] As early as 1907 and later in 1918, Störmer was able to show that orbits of a charged particle in the field of a magnetic dipole admit most remarkable periodic orbits. In a more recent publication, Zeits. f. Astrophys., 1, 247 (1930), Störmer describes two methods for the determination of periodic orbits, one of which led to the discovery of the family of unstable symmetric periodic orbits which forms part of this discussion (Cf. post section 3).

[5] V. Bush, J. Frank. Inst., 212, 447 (1931).

[6] G. Lemaître and M. S. Vallarta, Ann. de la Soc. Sci. de Bruxelles, A56, 102–103 (1936).

ceded by a brief summary of the properties of the differential equations governing the motion of charged particles in the field of a magnetic dipole, which were first set up and studied by Carl Störmer,[7] as early as 1904, and which have since been made generally well known through his numerous contributions.[8]

In terms of an independent variable σ and of the two dependent variables x, λ the differential equations of motion in the conformal transformation of the meridian plane take the form

$$\frac{d^2x}{d\sigma^2} = X(x, \lambda), \tag{1}$$

$$\frac{d^2\lambda}{d\sigma^2} = \Lambda(x, \lambda); \tag{2}$$

where X and Λ, hereafter to be referred to as the x and λ components of acceleration, are the functions of the coördinates x and λ specified below:

$$X(x, \lambda) = \frac{1}{16\gamma_1^4} e^{2x} - e^{-x} + e^{-2x}\cos^2\lambda, \tag{3}$$

$$\Lambda(x, \lambda) = e^{-2x}\sin\lambda\cos\lambda - \frac{\sin\lambda}{\cos^3\lambda}. \tag{4}$$

The present discussion refers to the study of orbits which originate asymptotically from a given periodic orbit and which form part of the boundary of the main cone, in the terminology adopted by Lemaître and Vallarta.[9] The computation of a family of asymptotic orbits for $\gamma_1 = 0.85$ comprises essentially these points: the calculation of the periodic orbit to which the required family of orbits is asymptotic; the determination of an asymptotic expansion representing accurately the desired family of orbits in the close neighborhood of the periodic orbit; and, finally, the numerical integration of the whole family of asymptotic

[7] Carl Störmer, "Sur le mouvement d'un point matériel portant une charge d'électricité sous l'action d'un aimant élémentaire," Vid. Selsk. Skr., Oslo (1904). For a more recent account of the equations of motion and of the *Störmer* unit of length, as regards the theory of cosmic radiation, see: Alfredo Baños, Jr., "Las ecuaciones del movimiento en la teoría de la radiación cósmica," *Ingeniería*, Vol. XI, No. 4 (1937).

[8] Carl Störmer, Zeits. f. Astrophys., 1, 237–274 (1930); references to his earlier work are given at the end of this paper. See also University Observatory, Oslo, Pub. No. 10 (1934).

[9] G. Lemaître and M. S. Vallarta, Phys. Rev., 50, 499 (1936).

orbits, by means of the methods developed recently by Lemaître and Vallarta and extended by the author in an effort to carry out the method as far away from the periodic orbit as seemed practicable.

3. **Calculation of the Periodic Orbit.** The calculation of the periodic orbit for $\gamma_1 = 0.85$ was effected by applying Lemaître's method.[10] This is an analytical extension of Störmer's empirical method, based on a complete analysis of all symmetric orbits cutting the x-axis ($\lambda = 0$) orthogonally, and affording a rapid, convenient and elegant way of determining, for a given value of γ_1, the outer member of the corresponding pair of symmetric periodic orbits. From the numerical verification of the periodic orbit, for $\gamma_1 = 0.85$, it is concluded that Lemaître's method yields an orbit which satisfies the differential equations of motion (1) and (2) correctly to four decimals. The principal limitation of the method arises from the use of Kepler's equation[11] to define the variation of $\sin \lambda$ as function of σ and which, being only an approximation, leads to a value of the period reliable only to four significant figures. For a complete account of Lemaître's method the reader is referred to the original article[10] and for its numerical application to $\gamma_1 = 0.93$ to a more recent paper by Lemaître and Vallarta.[12]

Lemaître's method giving the period $2\pi/\omega$, permits the computation of the x and λ coördinates of the periodic orbit for given values of $\omega\sigma$, say by increments of $\pi/12$ radians. From such a tabulated set of values it is then possible to compute, by applying well known methods of harmonic analysis,[13] the Fourier coefficients which exhibit the outer symmetric periodic orbit, in parametric representation, by means of the Fourier series:

$$x = z_0 + z_2 \cos 2(\omega\sigma + \varphi) + z_4 \cos 4(\omega\sigma + \varphi) + \cdots, \tag{5}$$

$$\lambda = \mu_1 \sin (\omega\sigma + \varphi) + \mu_3 \sin 3(\omega\sigma + \varphi) + \cdots. \tag{6}$$

To verify the precision with which (5) and (6) satisfy the differential equations of motion (1) and (2) it is then sufficient to compute directly from (5) and (6) the second order derivatives with respect to σ, and to equate the resulting Fourier series to the corresponding series for

[10] G. Lemaître, Ann. de la Soc. Sci. de Bruxelles, A54, 194–207 (1935).

[11] Ibid., p. 202.

[12] Loc. cit. ref. 6.

[13] E. T. Whittaker and G. Robinson, "The Calculus of Observations" (2nd ed.; London: Blackie and Son, 1932), p. 273.

the x and λ components of acceleration, obtained after replacing in (3) and (4), for x and λ, their corresponding Fourier series. To carry out rapidly the computation of the Fourier coefficients of the functions X and Λ, when x and λ are given by the Fourier series (5) and (6), as well as the Fourier coefficients corresponding to the higher order partial derivatives of X and Λ, the author devised schedules of computation following general prescriptions originally established by Lemaître and Vallarta.[14] A condensed description of the construction and application of these schedules of computation is given in the next section.

Such a comparison, as pointed out above, shows immediately the precision associated with the original periodic orbit as computed by means of Lemaître's method. To bring about improved agreement with the differential equations of motion corrections must be applied to the Fourier coefficients in (5) and (6) as well as to the parameter ω. These corrections are most conveniently obtained by making use of the variational equations,[15] which permit the study of periodic orbits in the infinitesimal neighborhood of the original uncorrected periodic orbit. To the final corrected periodic orbit for $\gamma_1 = 0.85$, in units of the 6th decimal, correspond the following values:

$z_0 =$	393,834.3	$\mu_1 =$	452,751.1
$z_2 =$	−12,011.8	$\mu_3 =$	−7,462.5
$z_4 =$	3.3	$\mu_5 =$	204.7
$\omega =$	916,331.7	$\mu_7 =$	−8.1
		$\mu_9 =$	0.4

4. **Harmonic Analysis of Fundamental Functions.** As stated in the preceding section, the verification of the precision associated with the periodic orbit, as computed by means of Lemaître's method, calls for the knowledge of the Fourier coefficients of the acceleration components X and Λ, as given in (3) and (4), when the dependent variables x and λ

[14] Op. cit., pp. 113–4.

[15] G. Lemaître and M. S. Vallarta, Ann. de la Soc. Sci. de Bruxelles, A56, 114 (1936). See also O. Godart, Ann. de la Soc. Sci. de Bruxelles, I, 58, 30 (1938). For a complete account of the derivation of the variational equations, as applied to the problem at hand, see Alfredo Baños, Jr., "Cálculo de las órbitas periódicas en la teoría de la radiación cósmica," forthcoming publication through *Ingeniería*, Mexico City (1939).

TABLE I

The Function $P(x, \lambda)$ and its Successive Partial Derivatives

Group	Derivative		Function	Fourier Series
I	P	P	$ae^{2x} + 2e^{-x} - e^{-2x}V - W$	$P(\sigma) = R_0 + R_2C_2 + R_4C_4 + \cdots$
II	X	$-\frac{1}{2}\frac{\partial P}{\partial x}$	$ae^{2x} - e^{-x} + e^{-2x}V$	$X(\sigma) = Z_0 + Z_2C_2 + Z_4C_4 + \cdots$
	$-\Lambda$	$-\frac{1}{2}\frac{\partial P}{\partial \lambda}$	$\frac{1}{2}e^{-2x}V' + \frac{1}{2}W'$	$-\Lambda(\sigma) = M_1S_1 + M_3S_3 + M_5S_5 + \cdots$
III	X_x	$\frac{1}{2}P_{xx}$	$2ae^{2x} + e^{-x} - 2e^{-2x}V$	$a(\sigma) = a_0 + a_2C_2 + a_4C_4 + \cdots$
	$-\Lambda_\lambda$	$-\frac{1}{2}P_{\lambda\lambda}$	$\frac{1}{2}e^{-2x}V'' + \frac{1}{2}W''$	$-b(\sigma) = b_0 + b_2C_2 + b_4C_4 + \cdots$
	$-X_\lambda = -\Lambda_x$	$-\frac{1}{2}P_{x\lambda}$	$-e^{-2x}V'$	$-c(\sigma) = c_1S_1 + c_3S_3 + c_5S_5 + \cdots$
IV	X_{xx}	$\frac{1}{2}P_{xxx}$	$4ae^{2x} - e^{-x} + 4e^{-2x}V$	$d(\sigma) = d_0 + d_2C_2 + d_4C_4 + \cdots$
	$X_{x\lambda} = \Lambda_{xx}$	$\frac{1}{2}P_{xx\lambda}$	$-2e^{-2x}V'$	$f(\sigma) = f_1s_1 + f_3S_3 + f_5S_5 + \cdots$
	$-X_{\lambda\lambda} = -\Lambda_{x\lambda}$	$-\frac{1}{2}P_{x\lambda\lambda}$	$-e^{-2x}V''$	$-g(\sigma) = g_0 + g_2C_2 + g_4C_4 + \cdots$
	$-\Lambda_{\lambda\lambda}$	$-\frac{1}{2}P_{\lambda\lambda\lambda}$	$\frac{1}{2}e^{-2x}V''' + \frac{1}{2}W'''$	$-h(\sigma) = h_1S_1 + h_3S_3 + h_5S_5 + \cdots$

$V = \cos^2\lambda$
$-\frac{1}{2}V'' = 2V - 1$
$-\frac{1}{2}V' = \sin\lambda\cos\lambda$
$\frac{1}{2}V''' = -2V'$

$W = \sec^2\lambda$
$\frac{1}{2}W'' = 3W^2 - 2W$
$\frac{1}{2}W' = \sin\lambda\sec^2\lambda$
$\frac{1}{2}W''' = 6WW' - 2W'$

are here replaced by their known Fourier expansions (5) and (6). This problem has already been treated by Lemaître and Vallarta who devised a method of harmonic analysis furnishing algebraic expressions for the required Fourier coefficients in terms of the known coefficients of the displacement components x and λ.

The method originally introduced by Lemaître and Vallarta[16] refers exclusively to the harmonic analysis of the functions X and Λ. In this work the author extended these new methods to the determination of the Fourier coefficients, not only of X and Λ, but also of the kinetic energy function P and its higher order partial derivatives up to the fourth order, as tabulated in *Table I*. Furthermore, to attain the desired accuracy in the computations corresponding to $\gamma_1 = 0.85$, to which the present discussion refers exclusively, the writer found it necessary to include, in each displacement component x and λ, two more harmonics than were employed by Lemaître and Vallarta in their computations for $\gamma_1 = 0.93$, and to retain, in the various expansions involved, higher order terms that had been originally neglected by Lemaître and Vallarta.

Table I gives a complete summary of the various functions studied together with the notation adopted to express their Fourier coefficients. Thus, for example, if the displacement components x and λ are known periodic functions of σ, as given in (5) and (6), then the kinetic energy function $P(x, \lambda)$, given in *Table I*, may be regarded as a periodic function of σ:

$$P(\sigma) = R_0 + R_2 \cos 2(\omega\sigma + \varphi) + R_4 \cos 4(\omega\sigma + \varphi) + \cdots$$

and the problem reduces to the computation of the R's from a previous knowledge of the z's and μ's. Evidently, it should always be possible to compute P for various values of σ and, hence, by well known methods of harmonic analysis, to compute the corresponding Fourier coefficients. This procedure, however, is extremely laborious and must be started anew whenever changes in the values of the z's and μ's are introduced. It was to obviate these shortcomings that Lemaître and Vallarta developed a direct method of harmonic analysis based on the fact that the functions involved possess known analytical expressions in terms of x and λ and that, consequently, it becomes possible to furnish approximately sufficiently accurate algebraic expressions for the required Fourier coefficients.

[16] G. Lemaître and M. S. Vallarta, Ann. de la Soc. Sci. de Bruxelles, A56, 108–114 (1936).

To extend the method of Lemaître and Vallarta to the harmonic analysis of the function P and its partial derivatives, as classified in four groups in *Table I*, it is first necessary to compute the Fourier coefficients of the individual functions of x and λ entering into the analytical expressions for P and its partial derivatives. These functions, as shown in column (3) of *Table I*, fall into three categories: 1) the exponentials e^{-x}, e^{-2x} and e^{2x}; 2) the trigonometric functions $\cos^2 \lambda$, $\sec^2 \lambda$ and their successive derivatives; 3) the product $e^{-2x} \cos^2 \lambda$ and its derivatives.

To simplify the discussion it will be convenient to introduce the symbolic notation:

$$S_k = \sin k(\omega\sigma + \varphi), \qquad C_k = \cos k(\omega\sigma + \varphi); \tag{7}$$

from which, after putting in evidence the fact that, in actual practice, the Fourier coefficients of (5) and (6) alternate in sign, these equations become:

$$x = z_0 \quad - z_2C_2 + z_4C_4 - \cdots = \sum_p (-1)^{p/2} z_p C_p, \tag{8}$$

$$\lambda = \mu_1 S_1 - \mu_3 S_3 + \mu_5 S_5 - \cdots = \sum_q (-1)^{\frac{q-1}{2}} \mu_q S_q; \tag{9}$$

where $p = 2k$ and $q = 2k + 1$, with $k = 0, 1, \cdots$ and where the z's and μ's are now all positive numbers.

A few examples will now be given. If we write

$$x = z_0 + \Delta x; \qquad \Delta x = -z_2C_2 + z_4C_4 - \cdots, \tag{10}$$

then, the problem at hand requires the computation of the Fourier coefficients in, say, the expansion of

$$e^{-2x} = A_0 + A_2C_2 + A_4C_4 + \cdots \tag{11}$$

from a knowledge of the z's, assuming that

$$e^{-2x} = e^{-2x_0}\{1 - 2\Delta x + 2(\Delta x)^2 - \tfrac{4}{3}(\Delta x)^3\}; \tag{12}$$

that is, assuming that $(\Delta x)^4$ is negligible. Following the method of Lemaître and Vallarta, it suffices to expand $(\Delta x)^2$ and $(\Delta x)^3$, as given in (10), reducing afterwards in the resulting expressions the products of sines and cosines, by means of well known trigonometric formulas, to the sum of harmonic components, which are finally collected to give the respective A's. The results are most conveniently arranged for

numerical work, as pointed out by Lemaître and Vallarta, in the form of schedules of computation, as shown below:

$$e^{-2x} = A_0 + A_2C_2 + A_4C_4 + \cdots$$

Factors	C_0	C_2	C_4	C_6
$\frac{1}{2}z_2^2e^{-2x_0}$		$3z_2$		z_2
$z_2e^{-2x_0}$	z_2	$-2z_4$	z_2	$-2z_4$
$2e^{-2x_0}$	$1/2$	z_2	$-z_4$	
$\Sigma\rightarrow$	A_0	A_2	A_4	A_6

Check: $e^{-2x_0} = A_0 + A_2 + A_4 + \cdots$

$$x_0 = z_0 - z_2 + z_4 - \cdots$$

Here, the various coefficients A_0, A_2, $\cdots$ are obtained by taking the sum of the terms in the corresponding column, multiplied by the factors written in the first column; thus, for example:

$$A_0 = e^{-2x_0}(1 + z_2^2),$$

$$A_2 = e^{-2x_0}(2z_2 - 2z_2z_4 + z_2^3).$$

As before, in the actual development (6), of λ as a periodic function of σ, it is found that the coefficient μ_1 of the fundamental is very much larger than the coefficients μ_3, μ_5, $\cdots$ of the higher harmonics, and that harmonics higher than the ninth are again negligible in computations involving the seventh decimal place. Writing

$$\lambda = \lambda_0 + \Delta\lambda, \qquad \lambda_0 = \mu_1S_1; \tag{13}$$

$$\Delta\lambda = -\mu_3S_3 + \mu_5S_5 - \mu_7S_7 + \mu_9S_9, \tag{14}$$

the problem at hand requires first the construction, for the functions $\cos^2\lambda$ and $\sec^2\lambda$, of expansions similar to the exponential development (11), and then, from such expansions, the computation of the corresponding Fourier coefficients. The details of these algebraic manipulations have already been given by Lemaître and Vallarta and are,

therefore, omitted here. As an example, the schedule of computation for the function $W = \sec^2 \lambda$ follows:

$$W = W_0 - W_2C_2 + W_4C_4 - \cdots$$

Factors	C_0	C_2	C_4	C_6	C_8	C_{10}
w_0'	μ_3	$2\mu_5$		μ_3	$2\mu_5$	
w_2'	μ_5	μ_3	$\frac{1}{2}\mu_3 + \mu_5$	μ_5	$\frac{1}{2}\mu_3$	μ_5
w_4'		$\frac{1}{2}\mu_3$	μ_3			$\frac{1}{2}\mu_3$
μ_3						
w_1		μ_3	$\mu_3 + \mu_5$	$\mu_5 + \mu_7$	$\mu_7 + \mu_9$	μ_9
w_3	μ_3	μ_5	μ_7	μ_3	μ_5	μ_7
w_5	μ_5	μ_3			μ_3	μ_5
w_7			μ_3			μ_3
1	w_0	w_2	w_4	w_6	w_8	w_{10}
$\Rightarrow$	W_0	W_2	W_4	W_6	W_8	W_{10}

Check: $1 = W_0 - W_2 + W_4 - W_6 + \cdots$

As indicated above, once in possession of the coefficients in the harmonic analysis of the elements entering into the fundamental function P, and its higher order partial derivatives with respect to x and λ, as given in *Table I*, the required Fourier coefficients, defined in the last column of this table, are simply constructed by forming the corresponding linear combinations specified in the preceding column. In this manner it was possible to effect the harmonic analysis of the fundamental functions required in these calculations. Thus, the knowledge of the Fourier coefficients corresponding to the functions P, X, and Λ permitted the correction of the periodic orbit to the desired precision, as pointed out in the preceding section. Likewise, the knowledge of the Fourier components of the higher order partial derivatives in Groups III and IV (Table I) served, respectively, for the computation of the first and second order variations of the periodic orbit, which are

essential to the computation of the characteristic exponent and of the asymptotic expansions giving the amplitudes of the pair of trigonometric series representing the x and λ coördinates of the asymptotic family of orbits in the immediate neighborhood of the periodic orbit.[17]

As explained in a later section, the coördinates of the asymptotic family of orbits may be expressed by means of the pair of trigonometric series:

$$x = z_0 - \sum_p (-1)^{\frac{p}{2}} (y_p S_p - z_p C_p), \tag{15}$$

$$\lambda = \sum_q (-1)^{\frac{q-1}{2}} (\mu_q S_q - \nu_q C_q), \tag{16}$$

where the coefficients y_p, z_p, μ_q and ν_q are to be regarded now as functions of the independent variable σ. When this is the case, the corresponding acceleration components assume the form of trigonometric series:

$$X = Z_0 + \sum_p (Y_p S_p + Z_p C_p), \tag{17}$$

$$\Lambda = \sum_q (M_q S_q + N_q C_q), \tag{18}$$

and the problem at hand now requires the computation of the Fourier coefficients Y_p, Z_p, M_q and N_q, from a knowledge of the Fourier expansions for the displacement components x and λ, as defined in equations (15) and (16).

To extend the methods sketched in the preceding paragraphs to the computation of the Fourier coefficients of the functions in Group II (Table I), when the displacement components are given by (15) and (16), where the coefficient ν_1 need not be small, it proves convenient to make an adequate change of phase, first introduced by Lemaître and Vallarta, which makes this coefficient ν_1 disappear.

Thus, if $\psi = (\omega\sigma + \varphi)$ denotes the phase associated with the trigonometric series (15) and (16); then, writing

$$\psi = \bar{\psi} + \psi_0, \tag{19}$$

and adopting the notation:

$$\begin{aligned} \bar{S}_k &= \sin k\bar{\psi}, \qquad \bar{C}_k = \cos k\bar{\psi}; \\ S_k^0 &= \sin k\psi_0, \qquad C_k^0 = \cos k\psi_0; \end{aligned} \tag{20}$$

[17] G. Lemaître and M. S. Vallarta, loc. cit., 115–120; see also Phys. Rev., 50, 497 (1936).

it may be readily verified that the expressions

$$\begin{aligned} \bar{y}_p &= z_p S_p^0 + y_p C_p^0 , \\ \bar{z}_p &= z_p C_p^0 - y_p S_p^0 ; \\ \bar{\mu}_q &= \mu_q C_q^0 + \nu_q S_q^0 , \\ \bar{\nu}_q &= \mu_q S_q^0 - \nu_q C_q^0 ; \end{aligned} \tag{21}$$

represent, in terms of $\bar{\psi} = (\omega\sigma + \varphi - \psi_0)$, the new amplitudes for x and λ in the expansions

$$x = z_0 - \sum_p (-1)^{\frac{p}{2}} (\bar{y}_p \bar{S}_p - \bar{z}_p \bar{C}_p), \tag{22}$$

$$\lambda = \sum_q (-1)^{\frac{q-1}{2}} (\mu_q S_q + \nu_q C_q), \tag{23}$$

which now replace (15) and (16). In particular, to make $\bar{\nu}_1$ vanish in (23) it suffices to determine ψ_0 adequately; for the purpose, putting $q = 1$ and $\bar{\nu}_1 = 0$ in the last equation (21), it follows that

$$\tan \psi_0 = \frac{\nu_1}{\mu_1}, \tag{24}$$

from which S_k^0, $C_k^0 (k = 1, 2, \cdots)$ are readily computed.

The trigonometric series of the acceleration components X and Λ, when expressed in terms of the modified phase $\bar{\psi} = (\omega\sigma + \varphi - \psi_0)$ and corresponding to the displacement components x and λ given in (22) and (23), may be written:

$$X = Z_0 + \sum_{p=2}^{6} (\bar{Y}_p \bar{S}_p + \bar{Z}_p \bar{C}_p), \tag{25}$$

$$\Lambda = \sum_{q=1}^{5} (\bar{M}_q \bar{S}_q + \bar{N}_q \bar{C}_q); \tag{26}$$

where the amplitudes $\bar{Y}_p$, $\bar{Z}_p$, $\bar{M}_q$, and $\bar{N}_q$ are computed, from a knowledge of the amplitudes in the trigonometric series (22) and (23), by applying the scheme of computation given in *Schedule* (*22*), reproduced here in Fig. 1. This schedule, designed to furnish the most compact arrangement for the sake of convenient calculations, was constructed by following the same principles outlined already and by following the prescriptions originally established by Lemaître and Vallarta; the details of this laborious piece of algebraic manipulation are, therefore, omitted. The reader is reminded to compare *Schedule*

$a =$	$\xi' - \frac{1}{2}\beta_1^2 \xi_1^2 =$			$\bar{x}_p = x_p C_1^\circ - y_p S_1^\circ$
$e^{\lambda_1 t} =$	$\xi_1' - \xi_1 \xi_2 = \xi_1 \bar{\xi}_2 =$	$\bar{\xi}_1' - \bar{\xi}_1 \bar{\xi}_2 =$		$\bar{y}_p = x_p S_1^\circ + y_p C_1^\circ$
$e^{\lambda_2 t} =$	$\xi_2' - \frac{1}{2}\beta_2^2 \xi_2^2 =$	$\bar{\xi}_2' - \bar{\xi}_2 \bar{\xi}_2 =$		$Z_p = \bar{Z}_p C_2^\circ - \bar{Y}_p S_2^\circ$
$e^{\lambda_3 t} =$	$\xi_3' - \xi_3 \xi_4 = \bar{\xi}_3 \bar{\xi}_4 =$	$\lambda_3' - \xi_3 \bar{\xi}_4 \xi_4 =$		$Y_p = \bar{Z}_p S_2^\circ + \bar{Y}_p C_2^\circ$

$\log e^{\lambda_4 t} =$ .....

	$\eta_1' - \frac{1}{2}\beta_1^2 \eta_1^2 =$	$\frac{p_1}{p_2} = \tan \psi_1 =$		$\bar{\mu}_2 = \mu_2 \xi_1^\circ + \mu_3 \zeta_2^\circ$
$\bar{\mu}_1' = \bar{\mu}_1 \eta_1 \eta_2 \bar{\eta}_3 =$	$\bar{\eta}_1' - \bar{\mu}_1 \bar{\eta}_1 \eta_2 \bar{\eta}_3 =$	$\sec^2 \psi_1 =$		$\bar{\nu}_2 = -\nu_2 \zeta_2^\circ + \mu_3 \xi_2^\circ$
$\bar{\mu}_2' = \bar{\mu}_2 \eta_3 =$	$\bar{\eta}_2' - \frac{1}{2}\beta_2^2 \eta_2^2 =$	$\sec \psi_2 =$		$M_2 = \bar{N}_2 \xi_2^\circ + \bar{N}_3 \zeta_2^\circ$
$\bar{\mu}_3' = \bar{\mu}_3 \eta_3 \eta_4 \bar{\eta}_4 =$	$\bar{\eta}_3' - \bar{\mu}_3 \bar{\eta}_3 \eta_4 \bar{\eta}_4 =$			$N_2 = \bar{N}_2 \zeta_2^\circ - \bar{N}_3 \xi_2^\circ$

FACTORS	C_0	C_2	C_4	C_6	S_2	S_4	S_6	FACTORS
Point No.	x_0	x_2	x_4	x_6	y_2	y_4	y_6	$\bar{x} = X$
	C_0°	C_2°	C_4°	C_6°	S_2°	S_4°	S_6°	
	$\bar{x}_0$	$\bar{x}_2$	$\bar{x}_4$	$\bar{x}_6$	$\bar{y}_2$	$\bar{y}_4$	$\bar{y}_6$	

$2e^{-\lambda_2 t}$	$\frac{1}{2} : 1$	$\bar{x}_2$	$-\bar{x}_4$	$\bar{x}_6$	$-\bar{y}_2$	$\bar{y}_4$	$-\bar{y}_6$	$e^{-\lambda_4 t}$
$2e^{-2\lambda_2 t}$	$\bar{x}_0'$	$-\bar{x}_2'$	$\bar{x}_4'$	$-\bar{x}_6'$	$\bar{y}_2'$	$-\bar{y}_4'$	$\bar{y}_6'$	$\frac{1}{2} e^{-\lambda_4 t}$
$\frac{1}{2}\beta_2 e^{\lambda_2 t}$		$2\bar{x}_0' + \bar{x}_4'$		$\bar{x}_4'$	$-\bar{y}_0'$		$-\bar{y}_4'$	$\frac{1}{2\beta_2} \bar{x}_2 e^{-\lambda_4 t}$
$\frac{1}{2}\beta_2 e^{2\lambda_2 t}$		$\bar{y}_0'$		$-\bar{y}_4'$	$-2\bar{x}_0' - \bar{x}_4'$		$-\bar{x}_4'$	$\frac{1}{2\beta_2} \bar{y}_2 e^{-\lambda_4 t}$
A, B	A_0	A_2	A_4	A_6	B_2	B_4	B_6	
	A_0'	A_2'	A_4'	A_6'	B_2'	B_4'	B_6'	$A_2' B_2'$

1	ν_0	ν_2	ν_4	ν_6				
ν_1		$\bar{\mu}_3$	$-\bar{\mu}_3 - \bar{\mu}_6$	$\bar{\mu}_6 + \bar{\mu}_9$	$-\bar{\nu}_3$	$\bar{\nu}_3 + \bar{\nu}_6$	$-\bar{\nu}_6 - \bar{\nu}_9$	
ν_3	$\bar{\mu}_3$	$-\bar{\mu}_6$	$\bar{\mu}_9$	$-\bar{\mu}_9$	$\bar{\nu}_6$	$-\bar{\nu}_9$	$\bar{\nu}_9$	
ν_5	$-\bar{\mu}_6$	$\bar{\mu}_9$			$\bar{\nu}_9$			
$2\nu_4 - 1$	$-\bar{\nu}_0'$	$\bar{\nu}_2'$		$\bar{\nu}_6'$	$-\bar{\mu}_3'$		$-\bar{\mu}_6'$	
ν_2	$\bar{\nu}_2'$	$-2\bar{\nu}_0'$	$\bar{\nu}_2' - \bar{\nu}_6'$	$-\bar{\nu}_0'$		$-\bar{\mu}_6' - \bar{\mu}_9'$	$\bar{\mu}_6'$	
ν_4		$\bar{\nu}_6'$	$-2\bar{\nu}_0'$		$-\bar{\mu}_6'$			
h_0	V_0	V_2	V_4	V_6	U_2	U_4	U_6	
$\frac{1}{2} h_2$	V_2	$2V_0 + V_4$	$V_2 + V_6$	V_4	U_4	$U_2 + U_6$	U_4	
$\frac{1}{2} h_4$	V_4	V_2	$2V_0$	V_2	$-U_2$		U_2	
$\frac{1}{2} h_6$			V_2	$2V_0$		$-U_2$		
$\frac{1}{2} k_2$	U_2	U_4	$-U_2 - U_6$	$-U_4$	$2V_0 - V_4$	$V_2 - V_6$	V_4	
$\frac{1}{2} k_4$	U_4	$U_2 - U_6$	U_4	U_2	V_2	$2V_0$	V_4	
$\frac{1}{2} k_6$		U_4	U_2		V_4	V_2	$2V_0$	
-1	A_0'	A_2'	A_4'	A_6'	B_2'	B_4'	B_6'	
$\frac{1}{2} e^{\lambda_2 t}$	A_2	$-A_2$	A_4	$-A_6$	$-B_2$	B_4	$-B_6$	
$\frac{1}{2\beta_2} e^{2\lambda_2 t}$		$-B_2'$	B_4	$-B_6'$	$\bar{y}_2'$	$-\bar{y}_4$	$\bar{y}_6'$	
Σ, Ψ	Z_0	Z_2	Z_4	Z_6	$\bar{Y}_2$	$\bar{Y}_4$	$\bar{Y}_6$	
Σ, Y	Z_0	Z_2	Z_4	Z_6	Y_2	Y_4	Y_6	

22

FACTORS	S_1	S_3	S_5	S_7	S_9	C_1	C_3	C_5	C_7	C_9
$\bar{\lambda} = \Lambda$	μ_1	μ_3	μ_5	μ_7	μ_9	ν_1	ν_3	ν_5	ν_7	ν_9
	S_1°	S_3°	S_5°	S_7°	S_9°	C_1°	C_3°	C_5°	C_7°	C_9°
	$\bar{\mu}_1$	$\bar{\mu}_3$	$\bar{\mu}_5$	$\bar{\mu}_7$	$\bar{\mu}_9$	$\bar{\nu}_1$	$\bar{\nu}_3$	$\bar{\nu}_5$	$\bar{\nu}_7$	$\bar{\nu}_9$

1	ν_1	ν_3	ν_5	ν_7						
$2\nu_4 - 1$		$-\bar{\mu}_3$	$\bar{\mu}_5 - \bar{\mu}_3$	$-\bar{\mu}_7$	$\bar{\mu}_9$		$-\bar{\nu}_3$	$\bar{\nu}_5$	$-\bar{\nu}_7$	$\bar{\nu}_9$
ν_2	$-\bar{\mu}_3$	$\bar{\mu}_5$	$-\bar{\mu}_5 - \bar{\mu}_7$	$\bar{\mu}_7 + \bar{\mu}_9$	$-\bar{\mu}_7$	$-\bar{\nu}_3$	$\bar{\nu}_5$	$-\bar{\nu}_5 - \bar{\nu}_7$	$\bar{\nu}_7 + \bar{\nu}_9$	$-\bar{\nu}_7$
ν_4	$\bar{\mu}_3 + \bar{\mu}_7$	$-\bar{\mu}_7$		$-\bar{\mu}_5$	$\bar{\mu}_5$	$-\bar{\nu}_3 + \bar{\nu}_7$	$-\bar{\nu}_7$		$-\bar{\nu}_7$	$\bar{\nu}_5$
ν_6	$-\bar{\mu}_5$	$\bar{\mu}_9$			$-\bar{\mu}_3$	$\bar{\nu}_5$	$-\bar{\nu}_9$			$-\bar{\nu}_9$
ν_1	$-2\bar{\nu}_0' - \bar{\nu}_4'$	$\bar{\nu}_4'$	$-\bar{\nu}_0'$	$\bar{\nu}_0' + \bar{\nu}_8'$	$-\bar{\nu}_8'$	$-\bar{\mu}_5'$	$\bar{\mu}_2'$	$-\bar{\mu}_0'$	$\bar{\mu}_0' - \bar{\mu}_8'$	$-\bar{\mu}_8'$
ν_3	$\bar{\nu}_2'$	$-2\bar{\nu}_0' - \bar{\nu}_8'$	$\bar{\nu}_4' - \bar{\nu}_8'$		$\bar{\nu}_8'$	$-\bar{\mu}_3'$	$-\bar{\mu}_0'$	$\bar{\mu}_2' + \bar{\mu}_8'$		$\bar{\mu}_8'$
ν_5	$-\bar{\nu}_0'$	$\bar{\nu}_2'$	$-2\bar{\nu}_0'$	$\bar{\nu}_2'$		$-\bar{\mu}_0'$	$-\bar{\mu}_2'$		$\bar{\mu}_2'$	
h_0	V_1	V_3	V_5	V_7	V_9	U_1	U_3	U_5	U_7	U_9
$\frac{1}{2} h_2$	$-V_1 + V_3$	$V_1 + V_5$	$V_3 + V_7$	$V_5 + V_9$	V_7	$U_1 + U_3$	$U_1 + U_5$	$U_3 + U_7$	$U_5 + U_9$	U_7
$\frac{1}{2} h_4$	$-V_3 + V_5$	$-V_1 + V_7$	V_1	V_3	V_5	$U_3 + U_7$	$U_1 + U_7$	U_1	U_3	U_5
$\frac{1}{2} h_6$	$-V_5$	$-V_3$	$-V_1$	V_1	V_3	U_5	U_3	U_1	U_1	U_3
$\frac{1}{2} k_2$	$U_1 - U_3$	$U_1 - U_5$	$U_3 - U_7$	$U_5 - U_9$	U_7	$V_1 + V_3$	$-V_1 + V_5$	$-V_3 + V_7$	$-V_5 + V_9$	$-V_7$
$\frac{1}{2} k_4$	$U_3 - U_5$	$U_1 - U_7$	U_1	U_3	U_5	$V_3 + V_5$	$V_1 + V_7$	$-V_1$	$-V_3$	$-V_5$
$\frac{1}{2} k_6$	U_5	U_3	U_1	U_1	U_3	V_5	V_3	V_1	$-V_1$	$-V_3$
1	$-m_1$	m_3	$-m_5$	m_7	$-m_9$					
$2m_4'$		$\bar{\mu}_3$	$-\bar{\mu}_5$	$\bar{\mu}_7$	$-\bar{\mu}_9$		$-\bar{\nu}_3$	$\bar{\nu}_5$	$-\bar{\nu}_7$	$\bar{\nu}_9$
m_2'	$-\bar{\mu}_3$	$\bar{\mu}_5$	$-\bar{\mu}_5 - \bar{\mu}_7$	$\bar{\mu}_7 + \bar{\mu}_9$	$-\bar{\mu}_7$	$-\bar{\nu}_3$	$\bar{\nu}_5$	$-\bar{\nu}_5 - \bar{\nu}_7$	$\bar{\nu}_7 + \bar{\nu}_9$	$-\bar{\nu}_7$
m_4'	$-\bar{\mu}_3 - \bar{\mu}_7$	$\bar{\mu}_7$		$\bar{\mu}_9$	$-\bar{\mu}_5$	$\bar{\nu}_3 - \bar{\nu}_7$	$-\bar{\nu}_7$		$\bar{\nu}_9$	$-\bar{\nu}_7$
m_6'	$-\bar{\mu}_5$	$\bar{\mu}_9$			$-\bar{\mu}_9$	$\bar{\nu}_5$	$-\bar{\nu}_9$			$-\bar{\nu}_9$
m_0'			$-\bar{\mu}_9$					$-\bar{\nu}_9$		
m_1'	$-2\bar{\nu}_0' - \bar{\nu}_4'$	$\bar{\nu}_4'$	$-\bar{\nu}_0'$	$\bar{\nu}_0' + \bar{\nu}_8'$	$-\bar{\nu}_8'$	$-\bar{\mu}_5'$	$\bar{\mu}_2'$	$-\bar{\mu}_0'$	$\bar{\mu}_0' + \bar{\mu}_8'$	$-\bar{\mu}_8'$
m_3'	$-\bar{\nu}_2'$	$2\bar{\nu}_0' + \bar{\nu}_8'$	$\bar{\nu}_4' - \bar{\nu}_8'$		$\bar{\nu}_8'$	$\bar{\mu}_3'$	$\bar{\mu}_0'$	$-\bar{\mu}_2' - \bar{\mu}_8'$		$\bar{\mu}_8'$
m_5'	$-\bar{\nu}_0'$	$\bar{\nu}_2'$	$-2\bar{\nu}_0'$	$\bar{\nu}_2'$		$-\bar{\mu}_0'$	$-\bar{\mu}_2'$		$\bar{\mu}_2'$	
M, N	M_1	M_3	M_5	M_7	M_9	N_1	N_3	N_5	N_7	N_9
M, N	M_1	M_3	M_5	M_7	M_9	N_1	N_3	N_5	N_7	N_9

FIG. 1

(*22*) with the corresponding pair of schedules given by Lemaître and Vallarta[18] observing that the requisite of higher precision and the necessity of including two more harmonics in the x and λ displacement components resulted in considerable more complication which, however, had to be endured in order to carry out successfully the method of numerical integration.

5. **Determination of the Asymptotic Expansion.** Once in possession of the periodic orbit, the study of the asymptotic family of orbits in its immediate vicinity is best carried out by integrating the variational equations derived from the differential equations of motion (1) and (2), using classical methods of integration developed by Poincaré and depending on the determination of the so-called *characteristic exponent* of the periodic orbit.[19]

This computation is most readily effected by postulating that a family of orbits asymptotic to a given periodic orbit, as given in parametric representation in (5) and (6), may be represented again in parametric form, by means of the two trigonometric series:

$$x = z_0(\sigma) + \sum_p \{y_p(\sigma) \sin p(\omega\sigma + \varphi) + z_p(\sigma) \cos p(\omega\sigma + \varphi)\}, \quad (27)$$

$$\lambda = \sum_q \{\mu_q(\sigma) \sin q(\omega\sigma + \varphi) + \nu_q(\sigma) \cos q(\omega\sigma + \varphi)\}, \quad (28)$$

where $p = 2k$ and $q = 2k + 1$, $(k = 0, 1, \cdots)$; and where the amplitudes $y_p(\sigma)$, $z_p(\sigma)$, $\mu_q(\sigma)$, and $\nu_q(\sigma)$ are now functions of σ, which, as σ approaches $-\infty$, must necessarily tend asymptotically to the corresponding amplitudes of the periodic orbit; that is, $z_p(\sigma)$ and $\mu_q(\sigma)$ tend to the corresponding Fourier coefficients z_p and μ_q of (5) and (6), and $y_p(\sigma)$, $\nu_q(\sigma)$ vanish asymptotically as σ goes to $-\infty$ since these coefficients are absent from the original representation of the periodic orbit. In equations (27) and (28) the arbitrary constant φ serves to distinguish individual members of the asymptotic family, thus permitting the computation of individual trajectories once the variation of the amplitudes as functions of σ is obtained for a given case.

To compute the variation of the amplitudes in the trigonometric series (27) and (28), in the immediate vicinity of the periodic orbit, recourse must be had to the classical method of characteristic expo-

[18] Op. cit., pp. 113–114.

[19] H. Poincaré, "Les Méthodes Nouvelles de la Mécanique Céleste" (Paris: Gauthier-Villars, 1892), Vol. I, Ch. IV; see also ref. (16), pp. 115–117, and ref. (3), p. 497.

nents developed by Poincaré, and recently applied to the problem at hand by Lemaître and Vallarta.[20] These computations furnish, in terms of $E = e^{\Omega\sigma}$, where Ω is the characteristic exponent of the periodic orbit, the asymptotic power series:

$$y_p(\sigma) = y_p + y'_p E + y''_p E^2 + \cdots, \tag{29}$$

$$z_p(\sigma) = z_p + z'_p E + z''_p E^2 + \cdots; \tag{30}$$

and similar expressions for $\mu_q(\sigma)$ and $\nu_q(\sigma)$, where the terms independent of σ correspond to the Fourier amplitudes of the periodic orbit as given in (5) and (6), and where the coefficients of E and E^2 represent, respectively, the first and second order terms obtained from the first and second variational equations of the periodic orbit. It will be recalled that the first variational equations furnish a simultaneous set of homogeneous equations in the first order coefficients y'_p, z'_p, μ'_q, and ν'_q, whose determinant must vanish thus yielding the "secular determinant," out of which it is possible to compute the smallest real positive root Ω; that is, the characteristic exponent of the periodic orbit.

The results of these computations, for $\gamma_1 = 0.85$, expressed in units of the sixth decimal, are summarized in the following table:

$$\Omega = 624{,}985$$

$z'_0 =$	$-500{,}000$	$\mu'_1 =$	$38{,}503$	$\mu''_1 =$	$159{,}451$
$z'_2 =$	$16{,}447$	$\nu'_1 =$	$-120{,}521$	$\nu''_1 =$	$-49{,}882$
$y'_2 =$	$4{,}287$	$\mu'_3 =$	$-5{,}535$	$\mu''_3 =$	-346
$z''_0 =$	$494{,}215$	$\nu'_3 =$	$2{,}854$	$\nu''_3 =$	$-6{,}965$
$z''_2 =$	$81{,}362$	$\mu'_5 =$	214	$\mu''_5 =$	$2{,}121$
$y''_2 =$	$15{,}472$	$\nu'_5 =$	-164	$\nu''_5 =$	79

6. The Lemaître-Vallarta Method of Numerical Integration. The knowledge of Ω and of the first and second order coefficients appearing in (29) and (30) allows the computation of the asymptotic variation of the amplitudes in the trigonometric series (27) and (28) in the immediate neighborhood of the periodic orbit. To investigate further this variation recourse must be had to a method of numerical integration devised by Lemaître and Vallarta.[21] For the purpose, it is first

[20] Loc. cit.

[21] G. Lemaître and M. S. Vallarta, Ann. de la Soc. Sci. de Bruxelles, A56, 120–130 (1936); see also Phys. Rev., 50, 497–499 (1936).

necessary to establish the differential equations governing the variation with σ of the amplitudes in the trigonometric series (27) and (28). This is readily attained from (1) and (2) by substituting, on the left hand side, the second order derivatives with respect to σ of the trigonometric series (27) and (28) and writing, for the right hand side functions of x and λ, the trigonometric series which are obtained by direct computation from (3) and (4), by replacing x and λ in these functions by their corresponding trigonometric series (27) and (28) and then using the methods originally developed by Lemaître and Vallarta as extended by the author to take care of the higher harmonics. The resulting set of simultaneous differential equations are finally obtained by equating the coefficients of corresponding harmonics in the phase $(\omega\sigma + \varphi)$. If we now represent the trigonometric series thus corresponding to (3) and (4) by means of the expressions

$$X = Z_0 + \sum_p \{Y_p(\sigma) \sin p(\omega\sigma + \varphi) + Z_p(\sigma) \cos p(\omega\sigma + \varphi)\}, \quad (31)$$

$$\Lambda = \sum_q \{M_q(\sigma) \sin q(\omega\sigma + \varphi) + N_q(\sigma) \cos q(\omega\sigma + \varphi)\}, \quad (32)$$

then, as shown by Lemaître and Vallarta, the resulting set of differential equations may be written:

$$\frac{d^2 y_p}{d\sigma^2} - 2p\omega \frac{dz_p}{d\sigma} - p^2\omega^2 y_p = Y_p, \quad (33)$$

$$\frac{d^2 z_p}{d\sigma^2} + 2p\omega \frac{dy_p}{d\sigma} - p^2\omega^2 z_p = Z_p; \quad (34)$$

where $p = 0, 2, \cdots$, with a similar set for μ_q, ν_q $(q = 1, 3, \cdots)$.

It will now be recalled that, for a given value of σ, the right hand terms Y_p, Z_p, M_q, and N_q may be evaluated, using the methods outlined in Section 4, directly in terms of the corresponding values of y_p, z_p, μ_q, and ν_q and that, accordingly, this lends itself readily to the point by point computation characteristic of all methods of numerical integration. The method originally employed by Lemaître and Vallarta has been extended by the author to include the use of higher harmonics through the computation of the additional coefficients of integration and the construction of a chart giving the scheme of computation to evaluate, for a given value of σ, the acceleration amplitudes Y_p, Z_p, M_q, and N_q in terms of the values for the corresponding displacement amplitudes (see *Schedule* (*22*) given in Fig. 1); and finally, its range of application has been considerably enlarged, to the full extent of the practical possibilities of the method, through the

derivation and application by the author of the formulas of numerical integration to effect the doubling and bisecting of the interval of integration. A thorough description of these details, together with a complete analysis of the Lemaître-Vallarta method of numerical integration is reserved, by the writer, for a forthcoming publication.[22]

7. **Results of the Numerical Integration.** The asymptotic power series (29) and (30) in the displacement amplitudes of the trigonometric series (27) and (28) allow the computation of the necessary steps to construct the table of successive differences required to initiate the method of numerical integration. Since the independent variable σ appears in the phase $(\omega\sigma + \varphi)$ multiplied by the factor ω, it proved convenient to choose for the interval of integration $\Delta\sigma$ the value defined thus:

$$\omega\Delta\sigma = \frac{\pi}{12} = 0.261,799; \qquad \Delta\sigma = 0.285,704$$

which represents an interval of integration considerably larger than the one usually employed by Störmer.[23]

The method employed to establish the initial system of successive differences has already been described by Lemaître and Vallarta.[24] It will only be mentioned here that the precision of tabulated values in the acceleration amplitudes Y_p, Z_p, M_q, and N_q, in so far as the equations of motion are concerned, indicates the fact that the initial system of successive differences in this case, for $\gamma_1 = 0.85$, is correct in the sixth decimal. The high precision thus attained may be regarded as an independent check on the computations of the characteristic exponent and of the coefficients of the first and second order terms in the asymptotic expansions (29) and (30).

The actual process of numerical integration turned out to be in this case an extremely laborious piece of work, comprising in all 54 steps, and requiring a total of approximately 500 computing hours, carried out according to the following program:

a) From $n = -12$ to $n = 12$, by double steps;
b) From $n = 12$ to $n = 26$, by single steps;
c) From $n = 26$ to $n = 35$, by half steps;
d) From $n = 35$ to $n = 37$, by extrapolation in half steps.

[22] Paper entitled: "On the Lemaître-Vallarta Method of Numerical Integration."

[23] Carl Störmer, Astrophysica Norvegica, 1, No. 4, p. 153 (1935). The interval of integration used by Störmer in this example is $\Delta\sigma = 0.015,625$.

[24] Op. cit., p. 120.

In carrying out the process of numerical integration it became necessary to consider additional higher harmonics in the displacement amplitudes in order to retain throughout the necessary accuracy. Furthermore, in an effort to extend the usefulness of the formulas of numerical integration it proved convenient to use fifth order differences. By watching the behavior of the last set of non-vanishing differences, in displacement and acceleration, the writer was able to estimate the accuracy of the method of numerical integration as it progressed and, in fact, this behavior served as a criterion for bisecting the double interval at $n = 12$ and the single interval at $n = 26$, for in the previous vicinity of these points the higher order differences began to grow rapidly. It may be affirmed, then, that sixth place accuracy was attained in the displacement amplitudes $y_p(\sigma)$, $z_p(\sigma)$, $\mu_q(\sigma)$, and $\nu_q(\sigma)$ from $n = -12$ to $n = 34$, except in the immediate vicinity preceding the bisecting of the interval at points 12, 26, and also 34, where the sixth decimal may be in error by two or three units.

The results of the numerical integration are summarized graphically in Fig. 2 which gives the variation, as functions of n (the ordinal number corresponding to each step in the process of numerical integration),[25] of the displacement and acceleration amplitudes appearing in the trigonometric series (27), (28), (31), and (32). The similarity of the general behavior for the 2nd, 4th and 6th harmonics corresponding to the x-components of displacement (full lines) and acceleration (dotted lines) is quite apparent in Fig. 2. Likewise, in the remaining group of figures, corresponding to the λ-components, the similarity in the general behavior of the 3rd, 5th, 7th, and 9th harmonics is evidently more striking, even to the extent of allowing the prediction that, for example, if the method of numerical integration had been continued further, the curves corresponding to the 3rd harmonic would eventually pass through the same maxima and minima which are already in evidence in the last set of curves corresponding to the 9th harmonic. The curves in Fig. 2 corresponding respectively to $p = 0$ and $q = 1$ and representing, therefore, the behavior of the most important terms in the trigonometric series for the x and λ components of displacement and acceleration, exhibit a variation not quite of identical fashion as their corresponding higher harmonics. The behavior of all curves, however, may be described in general by saying that the various

[25] The independent variable is of course σ, but it proved more convenient to retain n as abscissa, n being the number of the step in the process of numerical integration corresponding to increments of $\pi/12$ radians in $\omega\sigma$.

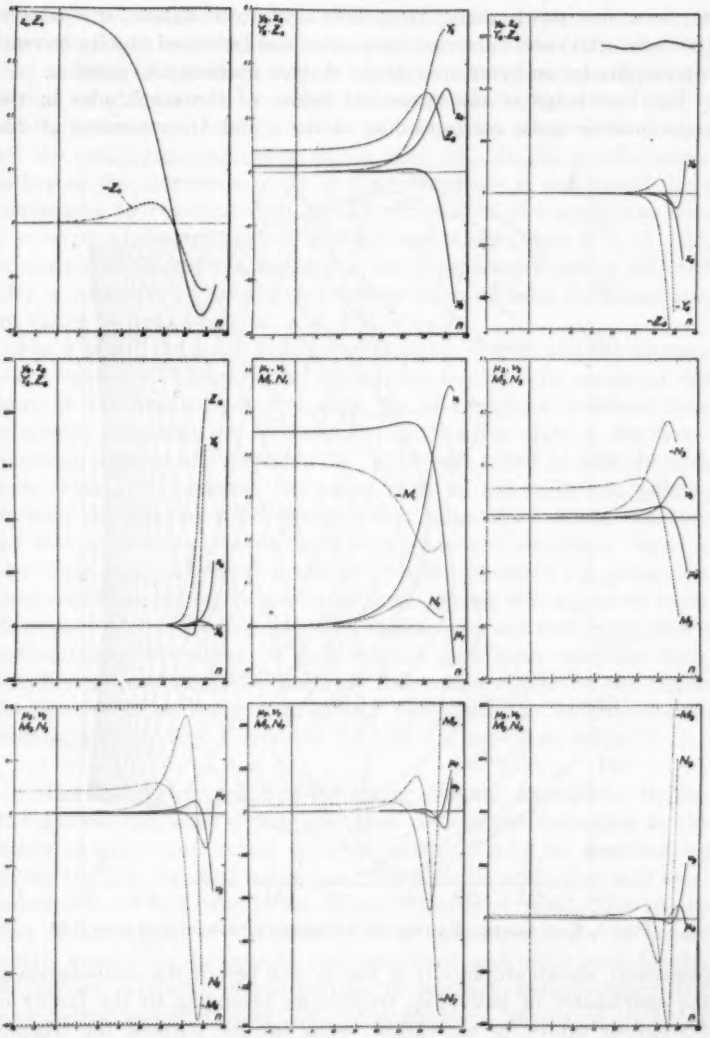


FIG. 2. RESULTS OF THE NUMERICAL INTEGRATION FOR $\gamma_1 = 0.85$ SHOWING THE DISPLACEMENT AND ACCELERATION AMPLITUDES AS FUNCTIONS OF THE INDEPENDENT VARIABLE.

curves at first slowly depart from their asymptotic values for $\sigma = -\infty$ (periodic orbit) and later start to execute oscillations of rapidly increasing amplitudes and correspondingly sharper maxima and minima.

The knowledge of the numerical values of the amplitudes in the trigonometric series corresponding to the x and λ components of dis-

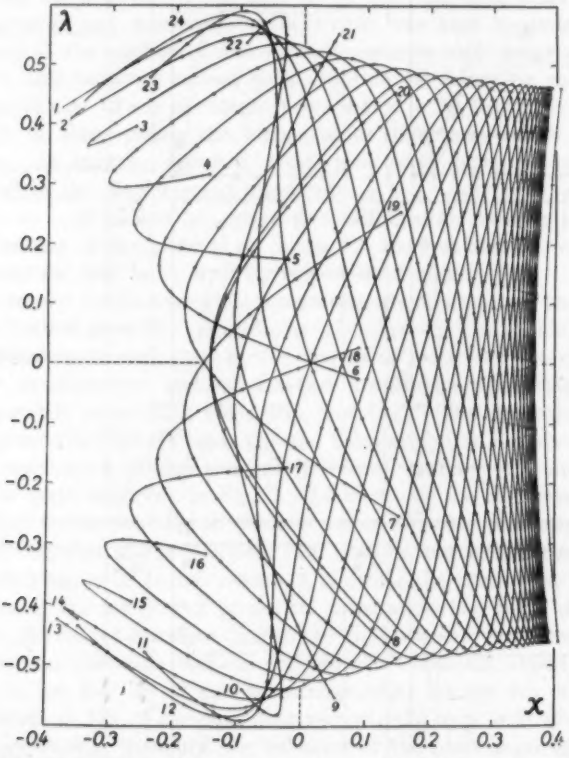


FIG. 3. CALCULATED FAMILY OF ASYMPTOTIC ORBITS FOR $\gamma_1 = 0.85$

placement, shown graphically in Fig. 2, now permit the computation of the coördinates of individual trajectories belonging to the family of asymptotic orbits for $\gamma_1 = 0.85$, using for the purpose the trigonometric series (27) and (28) and selecting the required orbits by assigning, in each case, suitable values of φ in the total phase $\psi = (\omega\sigma + \varphi)$.

To single out 24 individual trajectories evenly spaced it is necessary

to assign to φ, in units of $\pi/12$ radians, the values $\varphi = 0, 1, \cdots, 23$. It is observed, however, from the trigonometric series (27) and (28) that a change of π radians in $\psi = (\omega\sigma + \varphi)$, arising exclusively from a change in φ, leaves the x-coördinate intact and changes the sign of the λ-coördinate. Therefore, of the 24 orbits thus selected, 12 of them are the symmetric counterpart of the other 12. In the actual computation of the ephemeris of 24 asymptotic orbits it was found more convenient to compute first the coördinates of the equiphase lines ψ = constant, by assigning to ψ in succession the values $0, 1, \cdots, 11$, in units of $\pi/12$ radians, and using the trigonometric series (27) and (28) to compute the coördinates corresponding to each equiphase line, for values of n ranging from $n = 0$ to $n = 37$.

Fig. 3 shows the outer periodic orbit for $\gamma_1 = 0.85$ and the 24 computed orbits belonging to the asymptotic family. To construct the figure it was first necessary to draw the 12 equiphase lines and their symmetric counterparts, obtained by plotting for every x the corresponding positive and negative λ's. Care was taken to identify each point with two numbers: the value of ψ in the equiphase passing through the point and the corresponding value of n. In this manner the task of drawing a given orbit was materially simplified, for it is now clear that individual orbits are readily obtained by going progressively from one equiphase to the next, joining with a smooth curve those points for which the value of n increases by one unit every time a new equiphase is reached. Fig. 3 does not show these equiphase lines, but their general character is clearly discernible at least in the region between $x = 0$ and the periodic orbit. The values of φ, in units of $\pi/12$ radians, given in Fig. 3 serve to identify the individual orbits.

8. **The Symmetric and Self-Reversing Doubly Asymptotic Orbits.** The general character of the individual trajectories belonging to the family of asymptotic orbits depicted in Fig. 3 may be described by saying that, in breaking away from the periodic orbit, the individual trajectories at first execute an infinite number of oscillations in the close neighborhood of the periodic orbit and later begin to depart rapidly away from the periodic orbit and from each other until, finally, the various orbits no longer partake of common characteristics, but have rather acquired their individual behavior associated with the particular value of φ which serves to identify each member of the family of asymptotic orbits.

The existence of the simplest, symmetric, doubly asymptotic orbit,

known as the *Queen* orbit, which lies in Fig. 3 in the close neighborhood of the orbit $\varphi = 8$ and its symmetric companion $\varphi = 20$, may be predicted from general considerations on the theory of orbits. In fact, it may be shown in general that any orbit cutting the axis $\lambda = 0$ orthogonally must be symmetric with respect to the equator since the force components X and Λ appearing in the equations of motion (1) and (2) are respectively even and odd functions of λ. These considerations are sufficient to establish the existence of the symmetric, doubly asymptotic *Queen* orbit which occurs in Fig. 3 for a value of φ between 8 and 9, as may be readily ascertained by observing that orbit $\varphi = 8$ has a point with vertical tangent just above the line $\lambda = 0$ and $\varphi = 9$ just below.

The problem of actually computing the coördinates of the *Queen* orbit was, however, much more involved, since the value of $\varphi = \varphi_0$, which characterized the *Queen* orbit, had to be computed by interpolation. The method adopted by the writer to carry out this interpolation consisted essentially of determining the value of $dx/d\sigma$ at the equator, $\lambda = 0$, for a number of orbits corresponding to known values of φ in the neighborhood of the *Queen* orbit; and then determining, by inverse interpolation, the value of $\varphi = \varphi_0$ which would make $dx/d\sigma = 0$ at $\lambda = 0$. As a result of these computations the value of φ characterizing the Queen orbit turned out to be $\varphi_0 = 8.073{,}430$ in units of $\pi/12$ radians or $\varphi_0 = 121° 8' 47''$; that is, only $1° 8' 47''$ removed from its nearest neighbor $\varphi = 9$. From the nature of the interpolation involved in this case the value of φ_0 thus obtained is not reliable beyond the fourth decimal.

Once in possession of the value φ_0 the coördinates of the Queen orbit were immediately computed from the respective trigonometric series for values of n ranging from $n = 0$ to $n = 36$. The results of these computations are shown graphically in Fig. 4. Here, the symmetric character and the property of being doubly asymptotic; that is, asymptotic in the past as well as in the future, are strikingly brought out by the exactness with which the *Queen* orbit and its symmetric companion (shown in black dots and corresponding to $\varphi = \varphi_0 + \pi$) coincide with each other in the range from $n = 24$ to $n = 36$ where they overlap. To the accuracy of plotting in the drawing, which was somewhat better than the third decimal, the agreement is perfect throughout the overlapping portions.

General considerations on the theory of asymptotic orbits applied to the problem at hand indicate that, except for a denumerable set of

measure zero, which in this case includes the *Queen* orbit and the self-reversing, doubly asymptotic orbits discussed below, all orbits belonging to a given asymptotic family eventually cross the periodic orbit and go to infinity.[28] Therefore, it is concluded that, of the

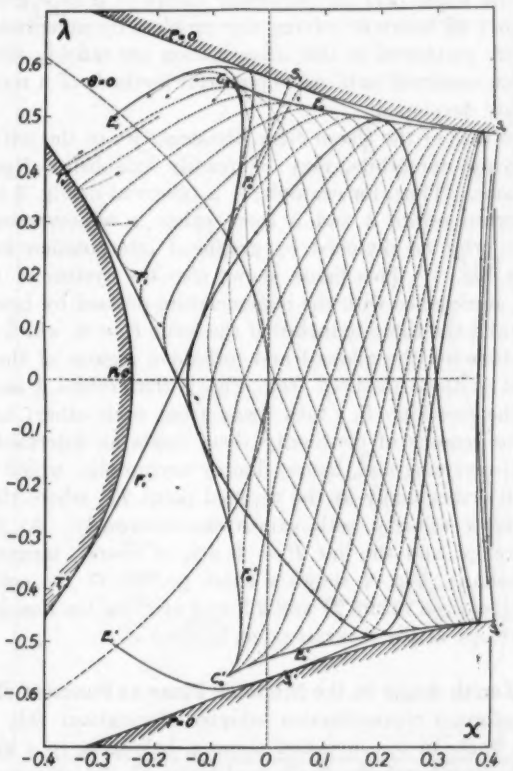


FIG. 4. THE QUEEN ORBIT, THE SELF-REVERSING DOUBLY ASYMPTOTIC ORBITS, AND THE SYSTEM OF ENVELOPES FOR $\gamma_1 = 0.85$.

trigonometric series (27) and (28), the one corresponding to the x-component of displacement yields a value of x which grows without limit as σ increases indefinitely, except for a denumerable set of measure zero in the assigned values of φ which serve to single out individual

[28] E. J. Schremp, Phys. Rev., 54, 153–157 (1938).

orbits and of which the particular value $\varphi = \varphi_0$, corresponding to the simplest, symmetric, doubly asymptotic *Queen* orbit, represents its most interesting example. This conclusion raises many difficult questions which arise when an attempt is made to examine the convergence of the trigonometric series (27) for particular values of φ as σ grows indefinitely. Since all hopes of solving this problem by numerical methods such as were employed in this investigation are quickly discarded, it must remain unsolved until more powerful methods of a more general character are developed.

As in the case of the *Queen* orbit, the existence of the self-reversing, doubly asymptotic orbits, may be readily established from general considerations. Thus, for example, it is observed in Fig. 3 that somewhere between orbits 2 and 3 there exists a self-reversing, doubly asymptotic orbit as sketched by graphical interpolation for a short portion, in Fig. 4. This figure shows also the system of envelopes, labelled in accordance with the nomenclature devised by Lemaître and Vallarta,[27] and the three branches of the locus $P = 0$, which serve as a boundary between the allowed and forbidden regions of the meridian plane.[28] It is then observed from Fig. 4 that orbits 2 and 3, after reaching the envelope F_1, turn away from each other; and, consequently, for reasons of continuity there exists an intermediate orbit exactly self-reversing and, hence, doubly asymptotic, which meets the line $P = 0$ orthogonally at the reversal point T_1, where the x and λ components of velocity both vanish simultaneously. At this point, the F_1 envelope and the line $P = 0$ are, of course, tangent to each other. Similarly, Fig. 4 shows a short portion of the self-reversing orbit lying between orbits 21 and 22, and meeting the line $P = 0$ and the E_1 envelope at the reversal point labelled S_1.

9. **The Zenith Angle in the Meridian Plane as Function of Latitude.** In the conformal representation adopted throughout this discussion the zenith angle in the meridian plane η, belonging to a given space trajectory, corresponds to the actual slope of the trajectory in the meridian plane as it crosses the locus $\sin \theta = 0$ (Fig. 4), where θ denotes the angle between the progressive tangent to the space trajectory and the meridian plane.[29] To investigate the variation of the zenith angle

[27] G. Lemaître and M. S. Vallarta, Phys. Rev., 49, 722-5 (1936).

[28] See, for example, Carl Störmer, University Observatory, Oslo, Pub. No. 10 (1934) and reference (7).

[29] For a thorough description of the angles θ and η, first introduced by Störmer, see article by the author mentioned in reference (7), or loc. cit. ref. 2.

in the meridian plane as a function of latitude it is required to measure, for each asymptotic trajectory intersecting the line $\theta = 0$, the angle η and the corresponding latitude λ. This was actually done for all trajectories intersecting the line $\theta = 0$ in the upper left hand quarter

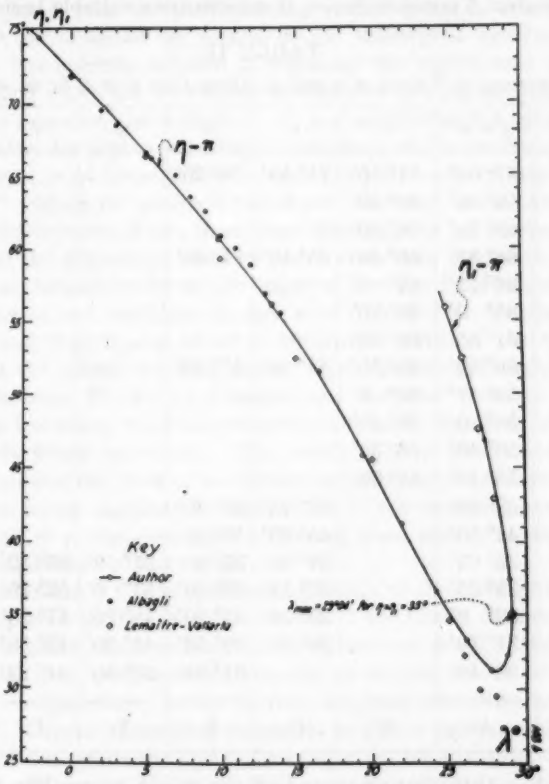


FIG. 5. THE ZENITH ANGLE IN THE MERIDIAN PLANE AS A FUNCTION OF LATITUDE.

of Fig. 3, which was redrawn for the purpose to a sufficiently large scale to guarantee the fourth decimal in the numerical results.

Fig. 5 gives the curves showing the variation, as functions of the latitude λ, of the angles η and η_1, corresponding to those sections of asymptotic orbits which intersect the line $\theta = 0$ coming respectively

from the envelopes E_0 and E_1 (Fig. 4). As shown in Fig. 5, the curves corresponding to η and η_1 are actually sections of a single smooth curve meeting at the point of maximum latitude $\lambda = 29° 4'$ for which $\eta = \eta_1 = 35°$. The $\eta - \lambda$ diagram of Fig. 5 is estimated to be correct to the nearest 5 minutes; hence, it constitutes a suitable basis of com-

TABLE II

Comparison of Values of η and η_1 Along Line $\theta = 0$ for $\gamma_1 = 0.85$

λ	$\eta - \pi$	η	η	η	η_1	η_1	$\eta_1 - \pi$
0°	72° 04′	71° 50′	71° 48′	74° 20′			
2°	69° 56′	69° 30′					
3°	68° 51′	68° 30′					
5°	66° 35′	66° 30′	65° 19′	71° 30′			
7°	64° 12′	65° 0′					
8°	63° 0′	63° 30′					
9°	61° 35′	62° 30′					
10°	60° 32′	60° 50′	59° 18′	67° 30′			
11°	59° 17′	60° 0′					
12°	58° 0′	58° 50′					
13°	56° 40′	56° 50′					
14°	55° 16′	55° 30′					
15°	53° 48′		52° 25′	60° 50′			
20°	44° 50′		45° 30′	50° 40′			
25°	35° 02′		34° 33′	35° 40′	57° 0′	55° 52′	55° 45′
26°	33° 25′		32° 12′	33° 30′	53° 0′	52° 26′	52° 20′
27°	32° 0′		29° 58′	31° 20′	48° 20′	47° 33′	48° 35′
28°	31° 10′		29° 31′	29° 50′	43° 30′	42° 58′	43° 50′
29°	33° 40′		—	31° 40′	33° 30′	34° 23′	36° 40′
Note:	A	B	C	D	D	C	A

Note A:—Author's values read off the $\eta - \lambda$ curve, Fig. 5.

Note B:—Albagli's values from her doctor's dissertation. (See Phys. Rev. 55, 15, 1939)

Note C:—Lemaître-Vallarta's critical angles for fundamental points as published in Phys. Rev., 50, p. 496, Table IV.

Note D:—Lemaître-Vallarta's values read off the corresponding curve for $\gamma_1 = 0.85$ in the $\eta - \lambda$ diagram, published in Phys. Rev. 49, p. 725, Fig. 5.

parison, for this particular value of γ_1, with the values of η, η_1 obtained by means of the differential analyzer. To facilitate this comparison corresponding values of η, η_1 have been tabulated, at properly chosen values of λ, in the columns labelled A of *Table II*. These values are to be compared with the corresponding entries given in columns B, C, and D, as obtained by means of the differential analyzer. In particular, the columns labelled D represent the values read off the corresponding curve for $\gamma_1 = 0.85$ in the $\eta - \lambda$ diagram published recently by Lemaître and Vallarta.[30] It is observed that the agreement is much better for higher latitudes in both the η and η_1 portions of the curve than it is at lower latitudes where the discrepancy may be as large as 7°. Since the author's values are estimated to be correct to the nearest 5 minutes of arc, these large discrepancies are attributed to the work of the differential analyzer which arise from systematic errors that are fully accounted for by the length of the runs involved, starting at the equator and reaching the line $\theta = 0$ after 3 or 4 units of σ. *Table II* and Fig. 5 also afford a comparison between the writer's values and the values found by means of the differential analyzer by Albagli[31] (column B) and by Lemaître and Vallarta[32] (column C), in both cases involving materially shorter runs and, therefore, yielding considerably better agreement. The results of these comparisons permit to affirm that the work of the differential analyzer yields asymptotic trajectories whose tangents are correct to 1° for runs not exceeding 3 to 4 units of σ, the error then increasing more rapidly to about 7° for runs longer than 10 units of σ.

10. **Conclusions.** 1. The outstanding result of these computations, already apparent in the complete drawing of the family of asymptotic orbits, is the proof of the existence of the symmetric, doubly asymptotic *Queen* orbit now unequivocally established to the extent to which numerical computations, correct to four decimals, constitute sufficient evidence. Without attempting at present to give a rigorous proof, the existence of the symmetric and self-reversing doubly asymptotic orbits leads to an important theorem, which may be regarded as the most important single result of the present investigations and which may be tentatively phrased as follows:

Theorem: The x-component of displacement described by the respec-

[30] Phys. Rev., 49, p. 725, Fig. 5.

[31] Reina Albagli, Ph.D. Thesis, Massachusetts Institute of Technology, December, 1937. See also Phys. Rev., 55, 15, 1939.

[32] G. Lemaître and M. S. Vallarta, Phys. Rev., 50, p. 496 (1936), Table IV.

tive trigonometric series, corresponding to a given family of asymptotic orbits, is unbounded as σ increases without limit, except for a denumerable set of measure zero in the values of φ which appear in the total phase $\psi = (\omega\sigma + \varphi)$.

2. The complete family of asymptotic orbits for $\gamma_1 = 0.85$, as drawn from the results of the numerical integration, permit a closer and more critical examination of the precision associated with the asymptotic trajectories drawn by means of the differential analyzer that had been possible heretofore. By comparing at a given latitude the zenith angles in the meridian plane for trajectories drawn by means of the differential analyzer with the corresponding zenith angles for the computed trajectories, it is concluded that the differential analyzer yields for short runs, not exceeding 3 to 4 units of σ trajectories whose inclinations in the conformal map are correct to better than 1°; and for longer runs, in excess of 10 units of σ, the error may increase to approximately 7°.

3. It is finally concluded, from these considerations, that the precision attained in drawing asymptotic trajectories by means of the differential analyzer, specially for shorter runs, is ample for all practical purposes connected with the theory of the main cone of allowed directions and allied topics such as the theory of the latitude and azimuthal effects of cosmic radiation.

11. **Acknowledgments.** Grateful appreciation is here extended to Professor M. S. Vallarta for his constant encouragement and able counsel throughout these investigations. For financial assistance, which afforded the necessary leisure to engage exclusively in the active prosecution of these studies, the writer is happy to acknowledge his debt of gratitude to the John Simon Guggenheim Memorial Foundation, for a fellowship extending over a period of 28 months; to the National University of Mexico and the Mexican Ministry of Education, for granting leave of absence and financial assistance, and to the Massachusetts Institute of Technology for two full-tuition scholarships.

Massachusetts Institute of Technology.

FLEXURE OF A RECTANGULAR COVERING SLAB ELASTICALLY FIXED BY ITS FOUR SIDES

By S. A. Savin

1. **Conditions of the problem.** The Theory of Structures considers three types of end fastenings of a member, or in other words, three degrees of restraint at the ends. Freely supported and hinged ends are the first type where the bending moment at the cross section over the support of the end under consideration, or shortly, its support moment is equal to zero, but the slope of the deflection curve, or what is the same thing, the deviation of this cross section is not equal to zero.

A completely or rigidly fixed end—such as fully restrained, built in, and clamped—is the second type where its support moment is not equal to zero, whereas its deviation is equal to zero.

A partially restrained or elastically fixed end is the third or intermedial type where both its support moment and deviation are different from zero. The third manner of end fixing of a member may be found in the Theory of Statically Indeterminate Constructions. A continuous beam and a framed construction, as known, are considered as a system of isolated beams to which the support moments are applied at their ends in addition to external forces (the given load and support reactions) directly acting on the span of a beam under consideration. Therefore, each beam of this system is a beam with both ends incompletely or elastically fixed.

The above three analogous types of fixing the sides of slabs may be encountered in a design of slabs of different forms covering a free space.

Let us consider, applying the methods of the mathematical theory of elasticity, a particular case of the flexure of a homogeneous isotropic covering slab of rectangular form with its elastically fixed sides.

Let us draw vertical plane sections through the sides of a rectangular covered area and call them the support sections of the slab covering this area. In consequence of the elastical fixing of the sides of this covering slab, bending moments and rotations at the points of its support sections will not vanish. A rectangular parallelepiped is formed from the intersection of four support sections with two external horizontal planes of this covering slab; this parallelepiped is the object of the discussions which follow. The four lower horizontal edges of

this parallelepiped are fixed in such a manner that all the points of these edges cannot have vertical displacements; besides, its four lower vertices are perfectly immovable.

The covering slab under consideration is bent under the influence of its own weight (q = the weight per unit of volume) and uniformly distributed load with intensity (p) over the upper horizontal plane and (p_0) over the lower one. In consequence of the uniformity of these vertical pressures and the homogeneity of the material of this covering slab, the rotations about the vertical axis must evidently be equal to zero at all the points of the covering slab.

2. **Rule of signs.** Let us place the origin of orthogonal coördinates in the centre of the covering slab, directing the vertical axis Z upwards and placing the horizontal axes X and Y in the median plane—the axis X across and Y along the covering slab—so that these axes form a right-handed system.

Let $2a$, $2b$, $2c$ denote the dimensions of the above rectangular parallelepiped along the axes X, Y, Z respectively.

It is known that a vector is determined by its magnitude, direction and sense. For all the vectors parallel to the axes of coördinates, as displacements, forces, stresses, we shall state the following general rule of signs: all such vectors shall be considered as positive when their senses coincide with the positive sense of the corresponding axis of coördinates and vice-versa. According to this rule of signs, the normal compressive unit stresses on the horizontal faces of the covering slab are equal to

$$\left.\begin{aligned} Z_z &= -p \text{ on the plane } z = c, \\ Z_{-z} &= p_0 \text{ on the plane } z = -c, \end{aligned}\right\} \tag{1}$$

where the lower indices (z) and ($-z$) show the sense of the outward normals to the upper and lower faces of the covering slab.

Thus according to this rule, the sign (+) corresponds to a tensile normal unit stress and the sign (−) to a compressive normal unit stress only when the sense of the outward normal to an elementary surface, taken parallel to any coördinate plane at the considered point of a body, coincides with the positive sense of the corresponding coördinate axis; if the sense of this outward normal coincides with the negative sense of the corresponding axis then the sign (+) gives a compressive and the sign (−) a tensile normal unit stress.

According to the adopted notations, the following active forces act

on the above mentioned parallelepiped, viz. its own weight $8abc \cdot q$ and the vertical pressures: $4ab \cdot p$ on the upper and $4ab \cdot p_0$ on the lower horizontal faces. According to the stated rule of signs, the first two forces are negative and the third positive.

Applying the mathematical terminology we shall name a point with the indefinite coördinates x, y, z the current point of a body.

The rotations in the current point of a body, according to Cauchy's formulas, can be determined from the system:

$$\frac{\omega_x}{\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}} = \frac{\omega_y}{\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}} = \frac{\omega_z}{\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}} = \frac{1}{2}, \tag{2}$$

where ω_x, ω_y, ω_z denote the rotations about the axes X, Y, Z and u, v, w the displacements along the same axes.

In accordance with these formulas, the positive directions of the rotations ω_x, ω_y, ω_z in every plane parallel to any coördinate plane are determined as indicated in the table below. According to the same rule of signs we shall determine the positive directions of the moments M_x, M_y, M_z of all forces acting on the forementioned parallelepiped with respect to axes X, Y, Z respectively.

Positive		In coördinate plane	Direction of rotation	
rotations	moments		from the first axis	to the second axis
ω_x	M_x	(Y, Z)	Y	Z
ω_y	M_y	(Z, X)	Z	X
ω_z	M_z	(X, Y)	X	Y

In this table it is supposed that the first coördinate axis in revolving from its original position to the second axis traces out one right angle.

3. **Expressions of displacements.** Applying his "Abridged formulas for the solution of differential equations of the theory of elasticity by means of integral polynomials" (Journal of Mathematics and Physics, vol. 17, number 4, Jan. 1939), the author has found for the covering slab under consideration the following expressions of displacements,— along the axis X

$$u = \frac{1+\sigma}{8 \cdot c^3} \frac{K}{E} [(2-\sigma)(c+z)^2(2c-z) - 3(1-\sigma)(b^2 - y^2)z]x, \tag{3}$$

along the axis Y

$$v = \frac{1+\sigma}{8 \cdot c^3} \frac{K}{E} [(2-\sigma)(c+z)^2(2c-z) - 3(1-\sigma)(a^2-x^2)z]y, \quad (4)$$

along the axis Z

$$w = -\frac{1+\sigma}{16 \cdot c^3} \frac{K}{E} [(1+\sigma)(c+z)^3(3c-z) + 3(1-\sigma)(a^2-x^2)(b^2-y^2) + 3\sigma(c^2-z^2)(a^2-x^2+b^2-y^2)] \quad (5)$$
$$+ \frac{(1+\sigma)(1-2\sigma)}{2(1-\sigma)} \frac{[q(c+z)-2p_0]}{E} (c+z)$$

introducing the abbreviated notation

$$K = 2qc - p_0 + p, \quad (6)$$

and denoting Young's modulus by E, Poisson's ratio by σ.

Let us now determine all the displacements at the points of the four lower horizontal edges of the parallelepiped under consideration. For the two transversal edges

$$z = -c, \qquad y = \pm b,$$

we have

$$w = 0, \qquad u = 0, \qquad v = \pm \frac{3}{8} \frac{1-\sigma^2}{c^2} \frac{K}{E} (a^2 - x^2)b,$$

and for the two longitudinal edges

$$z = -c, \qquad x = \pm a,$$

we obtain

$$v = 0, \qquad w = 0, \qquad u = \pm \frac{3}{8} \frac{1-\sigma^2}{c^2} \frac{K}{E} (b^2 - y^2)a.$$

Thus, in accordance with the elastical fixing of the support sections of the covering slab, all the four lower edges get some lateral displacements. However, the four lower vertices of the parallelepiped with the coördinates $(\pm a, \pm b, -c)$ remain perfectly immovable ($u = v = w = 0$), which corresponds to the initial conditions of the problem.

According to (5), the vertical deflection of the centre of the covering slab is

$$w_0 = -\frac{3}{16}\frac{1+\sigma}{c^3}\frac{K}{E}\{c^2[c^3+\sigma(a^2+b^2+c^2)]+(1-\sigma)a^2b^2\} + \frac{(1+\sigma)(1-2\sigma)}{2(1-\sigma)}\frac{qc-2p_0}{E}c. \tag{7}$$

4. **Cubical dilatation.** For obtaining cubical dilatation Δ we first find: according to (3) the transversal unit elongation along the axis X

$$\frac{\partial u}{\partial x} = \frac{1+\sigma}{8c^3}\frac{K}{E}[(2-\sigma)(c+z)^2(2c-z) - 3(1-\sigma)(b^2-y^2)z],$$

according to (4) the longitudinal unit elongation along the axis Y

$$\frac{\partial v}{\partial y} = \frac{1+\sigma}{8c^3}\frac{K}{E}[(2-\sigma)(c+z)^2(2c-z) - 3(1-\sigma)(a^2-x^2)z],$$

and according to (5) the vertical unit elongation along the axis Z

$$\frac{\partial w}{\partial z} = \frac{1+\sigma}{8c^3}\frac{K}{E}[3\sigma(a^2-x^2+b^2-y^2)z - 2(1+\sigma)(c+z)^2(2c-z)] + \frac{(1+\sigma)(1-2\sigma)}{1-\sigma}\frac{[q(c+z)-p_0]}{E}.$$

Summing up these three expressions we find the cubical dilatation

$$\Delta = \frac{(1+\sigma)(1-2\sigma)}{E}\left\{\frac{K}{8c^3}[2(c+z)^2(2c-z) - 3(a^2-x^2+b^2-y^2)z] + \frac{1}{1-\sigma}[q(c+z)-p_0]\right\}.$$

From the above formulas it is seen that the median plane ($z = 0$) of the covering slab obtains transversal and longitudinal expansions, for

$$\left[\frac{\partial u}{\partial x}\right]_{z=0} = \left[\frac{\partial v}{\partial y}\right]_{z=0} = \frac{(1+\sigma)(2-\sigma)}{4}\frac{K}{E}. \tag{8}$$

5. **Formulas of normal unit stresses.** For an elastic isotropic body the normal unit stresses X_x, Y_y, Z_z may—according to the theory of elasticity—be determined from the system

$$\frac{X_x}{\sigma\cdot\Delta + (1-2\sigma)\frac{\partial u}{\partial x}} = \frac{Y_y}{\sigma\cdot\Delta + (1-2\sigma)\frac{\partial v}{\partial y}} = \frac{Z_z}{\sigma\cdot\Delta + (1-2\sigma)\frac{\partial w}{\partial z}} = \frac{E}{(1+\sigma)(1-2\sigma)}.$$

Hence, according to the preceding formulas, the expressions of normal unit stresses are obtained as follows

$$X_x = \frac{K}{8c^3}\{(2+\sigma)(c+z)^2(2c-z) - 3z[\sigma(a^2-x^2)+b^2-y^2]\} + \frac{\sigma}{1-\sigma}[q(c+z)-p_0], \tag{9}$$

$$Y_y = \frac{K}{8c^3}\{(2+\sigma)(c+z)^2(2c-z) - 3z[a^2-x^2+\sigma(b^2-y^2)]\} + \frac{\sigma}{1-\sigma}[q(c+z)-p_0], \tag{10}$$

$$Z_z = -\frac{K}{4c^3}(c+z)^2(2c-z) + q(c+z) - p_0. \tag{11}$$

In these formulas K is determined by (6).

Hence, the normal unit stresses at all the points of the median plane ($z = 0$) are constant:

$$X_x = Y_y = \frac{(1+\sigma)(2-\sigma)}{2(1-\sigma)}qc - \frac{2+3\sigma-\sigma^2}{4(1-\sigma)}p_0 + \frac{2+\sigma}{4}p, \qquad Z_z = -\tfrac{1}{2}(p+p_0). \tag{12}$$

6. **Formulas of tangential unit stresses.** According to the formulas (3)–(5) the following expressions of partial derivatives are formed

$$\frac{\partial v}{\partial x} = \frac{\partial u}{\partial y} = \frac{3}{4}\frac{1-\sigma^2}{c^3}\frac{K}{E}xyz, \tag{13}$$

$$\frac{\dfrac{\partial w}{\partial y}}{(1-\sigma)(a^2-x^2)+\sigma(c^2-z^2)} = \frac{\dfrac{\partial v}{\partial z}}{(2-\sigma)(c^2-z^2)-(1-\sigma)(a^2-x^2)} = \frac{3}{8}\frac{1+\sigma}{c^3}\frac{K}{E}y, \cdots \tag{14}$$

$$\frac{\dfrac{\partial u}{\partial z}}{(2-\sigma)(c^2-z^2)-(1-\sigma)(b^2-y^2)} = \frac{\dfrac{\partial w}{\partial x}}{(1-\sigma)(b^2-y^2)+\sigma(c^2-z^2)} = \frac{3}{8}\frac{1+\sigma}{c^3}\frac{K}{E}x \cdots \tag{15}$$

It is seen from (13) that there is no shear deformation γ_z in the median plane ($z = 0$) as

$$\gamma_z = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = 0. \tag{16}$$

Tangential unit stresses

$$X_y = Y_x, \qquad Y_z = Z_y, \qquad Z_x = X_z, \tag{17}$$

as it is known, may be determined from the system

$$\frac{X_y}{\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}} = \frac{Y_z}{\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}} = \frac{Z_x}{\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}} = \frac{E}{2(1+\sigma)},$$

from which we find for the covering slab under consideration

$$\frac{X_y}{2(1-\sigma)xyz} = \frac{Y_z}{(c^2 - z^2)y} = \frac{Z_x}{(c^2 - z^2)x} = \frac{3}{8}\frac{K}{c^3}. \tag{18}$$

It is seen from these formulas that the tangential unit stresses X_z and Y_z on the horizontal planes ($z = \pm c$) of the covering slab are equal to zero which corresponds to the conditions of the problem.

For all the points of the median plane ($z = 0$), the formulas (18) give

$$X_y = 0, \qquad \frac{X_z}{x} = \frac{Y_z}{y} = \frac{3}{8}\frac{K}{c}. \tag{19}$$

Having the expressions (9)–(11) for the normal and (17), (18) for the tangential unit stresses, it is easy to verify that Navier's equations of the internal equilibrium

$$\left.\begin{aligned} \frac{\partial X_x}{\partial x} + \frac{\partial X_y}{\partial y} + \frac{\partial X_z}{\partial z} &= 0, \\ \frac{\partial Y_x}{\partial x} + \frac{\partial Y_y}{\partial y} + \frac{\partial Y_z}{\partial z} &= 0, \\ \frac{\partial Z_x}{\partial x} + \frac{\partial Z_y}{\partial y} + \frac{\partial Z_z}{\partial z} &= q, \end{aligned}\right\} \tag{20}$$

are satisfied identically.

7. **Notations and conditions for detailed verification of the equilibrium of the covering slab.** The two external horizontal faces of the covering slab together with the planes of its four support sections form, as it was noted above, a parallelepiped, the faces of which are characterized in the following table:

NN and names of the faces		Outward normal (n) to the face	Equation of the face
I	horizontal faces	$n = +Z$	$z = c$
II		$n = -Z$	$z = -c$
III	longitudinal sections	$n = +X$	$x = a$
IV		$n = -X$	$x = -a$
V	transversal sections	$n = +Y$	$y = b$
VI		$n = -Y$	$y = -b$

To verify the equilibrium of this parallelepiped it is necessary to determine the stresses acting on its faces, their resultants, and moments.

The resultants of the stresses parallel to the axes X, Y, Z we shall denote by S, T, V respectively, and compute separately for each face of the parallelepiped. The moments M_x, M_y, M_z of the stresses acting on each of the six faces of the parallelepiped we shall determine with respect to the rectangular axes X, Y, Z drawn from the centre of the covering slab.

If the sense of a force parallel to a coördinate axis (and consequently the sense of a unit stress) is unknown, we will suppose that this force acts along the positive sense of this coördinate axis. The moments of such forces with respect to an axis normal to the considered plane, which is parallel to one of the coördinate planes, we shall determine following the condition that the points of application of these forces have positive coördinates on the considered plane.

Forces as well as their moments and consequently unit stresses are vectorial quantities, therefore, the sign of the final result of the calculation of a sought quantity shall show its actual sense. Following these initial conditions we shall calculate the resultants of the normal and tangential stresses and their moments on the faces of the parallelepiped.

The unit stresses on all the faces of the parallelepiped are determined by the formulas of the theory of elasticity

$$\left.\begin{aligned} X_n &= X_x \operatorname{Cos}(x, n) + X_y \operatorname{Cos}(y, n) + X_z \operatorname{Cos}(z, n), \\ Y_n &= Y_x \operatorname{Cos}(x, n) + Y_y \operatorname{Cos}(y, n) + Y_z \operatorname{Cos}(z, n), \\ Z_n &= Z_x \operatorname{Cos}(x, n) + Z_y \operatorname{Cos}(y, n) + Z_z \operatorname{Cos}(z, n), \end{aligned}\right\} \quad \cdot \cdot \cdot \quad (21)$$

where (n) is an outward normal to the surface under consideration.

In accordance with Newton's principle of action and reaction, these formulas, as known, show that unit stresses at the same point acting across the opposite sides of the same surface are equal but act in opposite senses, because the outward normals to these sides of the surface have the opposite senses.

8. **Stresses, their resultants and moments on the I-st and II-nd faces of the parallelepiped.** According to the formulas (21), (17) and (18), the tangential unit stresses across the I-st and II-nd faces of the above parallelepiped are

$$(X_{\pm z})_{z=\pm c} = (Y_{\pm z})_{z=\pm c} = 0. \tag{22}$$

In these formulas the upper signs correspond to the I-st face and the lower ones—to the II-nd face.

The normal unit stresses, according to (21), (11), on the I-st and II-nd faces are respectively equal to

$$(Z_z)_{z=c} = -p, \qquad (Z_{-z})_{z=-c} = -(Z_z)_{z=-c} = p_0, \tag{23}$$

which, according to (1), corresponds to the conditions of the problem. The formulas (23) show, according to the stated rule of signs, that the covering slab is pressed from above and from below by the vertical uniformly distributed pressures p and p_0.

According to (22) the resultants of the tangential stresses parallel to the axes X, Y are respectively equal to

$$S = T = 0. \tag{24}$$

According to (23) the resultants of the normal stresses parallel to the axis Z on the I-st and II-nd faces are respectively equal to

$$V_1 = -4pab, \qquad V_2 = 4p_0ab. \tag{25}$$

In consequence of absence of the tangential stresses and the uniform distribution of the normal stresses, the moments of all the stresses acting on the I-st and II-nd faces about the axes X, Y, Z with origin at the centre of the covering slab are

$$M_x = M_y = M_z = 0. \tag{26}$$

9. **Stresses, their resultants and moments on the III-rd and IV-th faces of the parallelepiped.**

According to (21) and (9), the normal unit stresses on the III-rd and IV-th faces are

$$(X_{\pm x})_{x=\pm a} = \pm\frac{K}{8c^3}[(2+\sigma)(2c^3+3c^2z-z^3)-3(b^2-y^2)z] \pm \frac{\sigma}{1-\sigma}[q(c+z)-p_0]. \tag{27}$$

According to (21), (17) and (18) the tangential unit stresses across the same two faces are

$$(Y_{\pm x})_{x=\pm a} = \pm(\pm a)\cdot\tfrac{3}{4}(1-\sigma)K\frac{yz}{c^3} = \tfrac{3}{4}(1-\sigma)K\frac{ayz}{c^3}, \tag{28}$$

$$(Z_{\pm x})_{x=\pm a} = \pm(\pm a)\cdot\tfrac{3}{8}K\frac{c^2-z^2}{c^3} = \tfrac{3}{8}K\frac{(c^2-z^2)a}{c^3}. \tag{29}$$

The resultants of the stresses are

$$S = \int_{-b}^{b}dy\int_{-c}^{c}(X_{\pm x})_{x=\pm a}\cdot dz = \pm bc\left[(2+\sigma)K+\frac{4\cdot\sigma}{1-\sigma}(qc-p_0)\right], \tag{30}$$

$$T = \int_{-b}^{b}dy\int_{-c}^{c}(Y_{\pm x})_{x=\pm a}\cdot dz = 0, \tag{31}$$

$$V = \int_{-b}^{b}dy\int_{-c}^{c}(Z_{\pm x})_{x=\pm a}\cdot dz = Kab. \tag{32}$$

According to (27)–(29) the moments about the same axes of all the stresses acting on the III-rd and IV-th faces (fig. 1) are

$$M_x = \int_{-c}^{c}dz\int_{-b}^{b}(Z_{\pm x})_{x=\pm a}\cdot y\,dy - \int_{-b}^{b}dy\int_{-c}^{c}(Y_{\pm x})_{x=\pm a}\cdot z\,dz = 0, \tag{33}$$

$$M_y = \int_{-b}^{b}dy\int_{-c}^{c}(X_{\pm x})_{x=\pm a}\cdot z\,dz - (\pm a)\int_{-b}^{b}dy\int_{-c}^{c}(Z_{\pm x})_{x=\pm a}\cdot dz = \pm b\left\{K\left[\tfrac{2}{3}(2+\sigma)c^2-a^2-\frac{b^2}{3}\right]+\frac{4}{3}\frac{\sigma}{1-\sigma}qc^2\right\}, \tag{34}$$

$$M_z = (\pm a)\int_{-b}^{b}dy\int_{-c}^{c}(Y_{\pm x})_{x=\pm a}\cdot dz - \int_{-c}^{c}dz\int_{-b}^{b}(X_{\pm x})_{x=\pm a}\cdot y\,dy = 0. \tag{35}$$

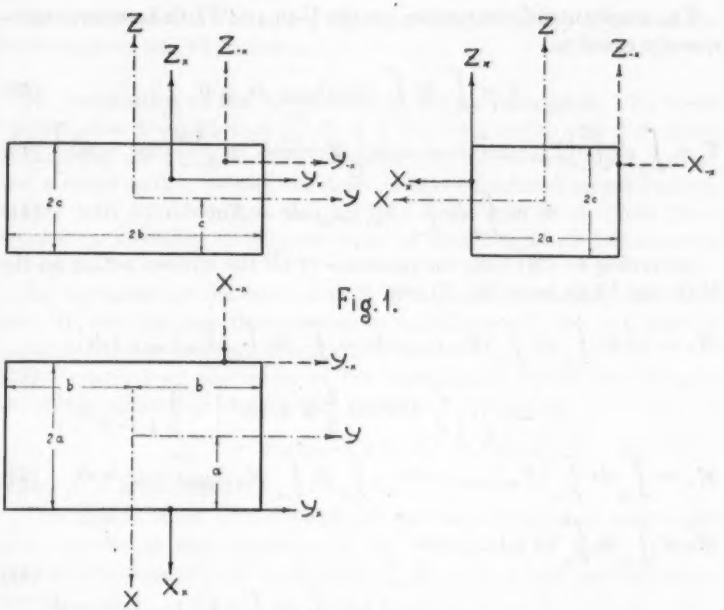


Fig. 1.

In all the formulas with double signs, the upper signs correspond to the III-rd face and the lower signs—to the IV-th face of the parallelepiped.

10. **Stresses, their resultants and moments on the V-th and VI-th faces of the parallelepiped.** According to the formulas (21), (17) and (18) the tangential unit stresses across the V-th and VI-th faces are

$$(Z_{\pm v})_{y=\pm b} = \pm(\pm b)\cdot\tfrac{3}{8}K\frac{c^2-z^2}{c^3} = \tfrac{3}{8}K\frac{c^2-z^2}{c^3}b \tag{36}$$

$$(X_{\pm v})_{y=\pm b} = \pm(\pm b)\cdot\tfrac{3}{4}K(1-\sigma)\frac{zx}{c^3} = \tfrac{3}{4}(1-\sigma)K\frac{zx}{c^3}b. \tag{37}$$

According to (21) and (10) the normal unit stresses on these two faces are

$$(Y_{\pm v})_{y=\pm b} = \pm\frac{K}{8\cdot c^3}[(2+\sigma)(2c^3+3c^2z-z^3)-3(a^2-x^2)z] \pm \frac{\sigma}{1-\sigma}[q(c+z)-p_0] \tag{38}$$

The resultants of the stresses on the V-th and VI-th faces are consequently equal to

$$S = \int_{-a}^{a} dx \int_{-c}^{c} (X_{\pm y})_{y=\pm b} \cdot dz = 0 \tag{39}$$

$$T = \int_{-a}^{a} dx \int_{-c}^{c} (Y_{\pm y})_{y=\pm b} \cdot dz = \pm ca \left[(2 + \sigma)K + \frac{4.\sigma}{1 - \sigma} (qc - p_0) \right] \tag{40}$$

$$V = \int_{-a}^{a} dx \int_{-c}^{c} (Z_{\pm y})_{y=\pm b} \cdot dz = Kab \cdots \tag{41}$$

According to (36)–(38) the moments of all the stresses acting on the V-th and VI-th faces (fig. 2) are

$$M_x = (\pm b) \int_{-a}^{a} dx \int_{-c}^{c} (Z_{\pm y})_{y=\pm b} \cdot dz - \int_{-a}^{a} dx \int_{-c}^{c} (Y_{\pm y})_{y=\pm b} \cdot z\, dz$$
$$= \pm a \left\{ K \left[\frac{a^2}{3} + b^2 - \frac{2}{5}(2 + \sigma)c^2 \right] - \frac{4}{3} \frac{\sigma}{1 - \sigma} qc^3 \right\} \tag{42}$$

$$M_y = \int_{-a}^{a} dx \int_{-c}^{c} (X_{\pm y})_{y=\pm b} \cdot z\, dz - \int_{-c}^{c} dz \int_{-a}^{a} (Z_{\pm y})_{y=\pm b} \cdot x\, dx = 0 \tag{43}$$

$$M_z = \int_{-c}^{c} dz \int_{-a}^{a} (Y_{\pm y})_{y=\pm b} \cdot x\, dx$$
$$- (\pm b) \int_{-c}^{c} dz \int_{-a}^{a} (X_{\pm y})_{y=\pm b} \cdot dx = 0 \tag{44}$$

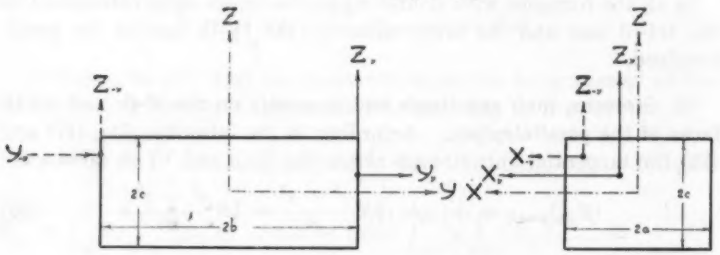
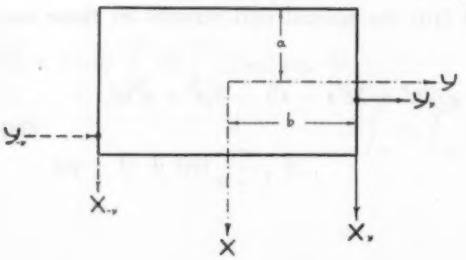


Fig. 2.



In these formulas the upper signs correspond to the V-th and the lower ones to the VI-th faces.

11. **Verification of the equilibrium of the parallelepiped.** To verify the equation of equilibrium $\sum X_k = 0$, it is sufficient to sum the values, obtained in the formulas (24), (30) and (39), of the resultants S of all the stresses acting on the six faces of the considered parallelepiped, taking into consideration that on the III-rd and IV-th faces these resultants, according to (30), are equal in their magnitude but opposite in sign.

By summing up the six values T given in the formulas (24), (31) and (40) we find that the equation of equilibrium $\sum Y_k = 0$ for this parallelepiped is satisfied identically.

By summing up the values of the resultants V for all the six faces we obtain, according to (25), (32) and (41)

$$\sum V = 4ab(p_0 - p + K) = 8abc \cdot q$$

where K is taken from (6).

This sum is equal to the weight of the parallelepiped in magnitude and opposite in sign according to the stated rule of signs. Consequently the equation of equilibrium $\sum Z_k = 0$ is also satisfied identically.

The moments of all the stresses on the six faces of the parallelepiped are determined with respect to the same axes X, Y, Z with the origin in the centre of the covering slab. Therefore—to verify: 1) the equation of equilibrium $\sum M_x = 0$, it will be sufficient to sum all the six values M_x according to (26), (33) and (42); 2) the equation $\sum M_y = 0$, it will be sufficient to sum all the six values M_y according to (26), (34) and (43); 3) the equation $\sum M_z = 0$, it will be sufficient to sum all the six values M_z according to (26), (35) and (44). Taking into consideration the double signs at (34) and (42) for the moments on the opposite faces, it is easy to be convinced that the three last equations are also satisfied identically.

The result of the detailed verification of the equilibrium by six equations might have been foreseen, as the obtained expressions of the unit stresses (9), (10), (11), (17), and (18) make the equations of the internal equilibrium (20) identical, as it was noticed in §6.

12. **Determination of the support moments.** The normal force, as well as the horizontal and vertical shear forces at both longitudinal

support sections are defined by the formulas (30), (31) and (32) respectively, and the corresponding forces at both transversal support sections are defined by (40), (39) and (41) respectively.

In order to define the support moments at these four support sections of the covering slab we first draw through the centre of each support section the new coördinate axes X_k, Y_k, Z_k ($k = 3, 4, 5, 6$) with the same senses parallelly to the old axes X, Y, Z respectively, and then we determine the moments M_x, M_y, M_z of all the stresses acting at the support section under consideration with respect to these new axes.

It is easy to be convinced that for both longitudinal sections ($x = \pm a$ with $n = \pm X$, $k = 3$ and 4) the expression (33) given for M_x does not become modified by the translation of the origin of coördinates. The unit stresses $Y_{\pm x}$ and $Z_{\pm x}$ act across the planes of the corresponding support sections, in which the new axes Y_k and Z_k ($k = 3$ and 4) are drawn; therefore the formulas (34) and (35) for the new axes are simplified (fig. 1):

$$M_y = \int_{-b}^{b}\int_{-c}^{c} (X_{\pm x})_{x=\pm a}\cdot z\,dy\,dz, \qquad M_z = -\int_{-b}^{b}\int_{-c}^{c} (X_{\pm x})_{x=\pm a}\cdot y\,dy\,dz$$

Thus, we find for both longitudinal support sections, according to (27), the expressions of the support moments

$$M_z = M_x = 0$$

$$M_y = \pm b\left\{K\left[\tfrac{2}{3}(2+\sigma)c^2 - \frac{b^2}{3}\right] + \frac{4}{3}\frac{\sigma}{1-\sigma}qc^3\right\}. \tag{45}$$

Here the upper signs correspond to the III-rd and the lower ones to the IV-th faces of the above parallelepiped.

Thus, at both longitudinal support sections the twisting moment, or torque, M_x as well as the bending moment M_z vanish whereas the bending moment M_y is not equal to zero.

For the two transversal support sections ($y = \pm b$ with $n = \pm Y$, $k = 5$ and 6), as it is seen (fig. 2), the expression M_y, being referred to the new axes, gives the same formula (43). Then the formulas (44) and (42) become simplified

$$M_z = \int_{-c}^{c}\int_{-a}^{a} (Y_{\pm y})_{y=\pm b}\cdot x\,dz\,dx, \qquad M_x = -\int_{-c}^{c}\int_{-a}^{a} (Y_{\pm y})_{y=\pm b}\cdot z\,dz\,dx.$$

Consequently we obtain, according to (38), for the transversal support sections of the covering slab the expressions of the support moments

$$M_x = \mp a\left\{K\left[\tfrac{1}{3}(2+\sigma)c^2 - \frac{a^2}{3}\right] + \frac{4}{3}\frac{\sigma}{1-\sigma}qc^2\right\}, \tag{46}$$

$$M_y = M_z = 0.$$

Here the upper signs correspond to the V-th and the lower to the VI-th faces of the same parallelepiped.

Consequently, at both transversal support sections the twisting moment M_y and the bending moment M_z vanish whereas the bending moment M_x is not equal to zero.

The bending moments (45) and (46), which do not vanish, correspond to the elastical fixing for the four support sections.

If the pressures p and p_0 do not vary then the support moments (45) and (46) may vanish but only in a square covering slab the dimensions of which are determined from the following condition

$$a = b = 2c\sqrt{0,3(2+\sigma) + \frac{\sigma}{1-\sigma}\frac{qc}{K}}, \tag{47}$$

where the constant K is computed by (6); in this case such a square covering slab will be simply supported but not elastically fixed.

13. **Rotations in points of the parallelepiped.** Using the formulas (2), according to (13)–(15), we find the expressions ω_x, ω_y, ω_z of the rotations at the current point of the covering slab:

$$\frac{\omega_x}{(a^2 - x^2 - c^2 + z^2)y} = \frac{\omega_y}{(c^2 - z^2 - b^2 + y^2)x} = \frac{3}{8}\frac{1-\sigma^2}{c^3}\frac{K}{E}, \tag{48}$$

$$\omega_z = 0. \tag{49}$$

The formula (49) shows that the rotations about the vertical axis Z are equal to zero at all the points of the covering slab, which corresponds to the conditions of the problem.

The eight vertices of the parallelepiped under consideration, as it is seen from the formulas (48), obtain no rotations about the axes X and Y.

For all the points of the upper transversal edges of the parallelepiped

$$z = c, \qquad y = \pm b,$$

and of lower ones

$$z = -c, \qquad y = \pm b,$$

the rotations, according to (48), are equal to

$$\omega_x = \pm\frac{3}{8}\frac{1-\sigma^2}{c^3}\frac{K}{E}b(a^2-x^2), \qquad \omega_y = 0. \tag{50}$$

For all the points of the four longitudinal edges

$$z = c, \qquad x = \pm a,$$
$$z = -c, \qquad x = \pm a,$$

according to (48), we obtain

$$\omega_x = 0, \qquad \omega_y = \mp\frac{3}{8}\frac{1-\sigma^2}{c^3}\frac{K}{E}a(b^2-y^2) \tag{51}$$

In all these formulas the double signs are mutually corresponding. The rotations ω_x in the I (50) and ω_y in the II (51), which do not vanish, correspond to the elastical fixing of the transversal and longitudinal edges of the parallelepiped respectively.

These rotations, evidently, also do not vanish in a square covering slab, the dimensions of which have been determined according to (47), and therefore all its support moments become annulled.

14. **Form of deformed median plane of the covering slab.** Deformations at the points of the median plane ($z = 0$) of the covering slab under consideration are defined by the formulas (7), (8) and (16), as well as the unit stresses by (12) and (19).

After deforming the orthogonal coördinates ξ, η, ζ of a current point of this median plane are expressed, according to (3)–(5), by the formulas:

$$\xi = x + u = x\left[1 + \tfrac{1}{4}(1+\sigma)(2-\sigma)\frac{K}{E}\right],$$

$$\eta = y + v = y\left[1 + \tfrac{1}{4}(1+\sigma)(2-\sigma)\frac{K}{E}\right],$$

$$\zeta = z + w = -\frac{3}{16}\frac{1+\sigma}{c^3}\frac{K}{E}[(1+\sigma)c^4 + (1-\sigma)(a^2-x^2)(b^2-y^2)$$
$$+ \sigma(a^2 - x^2 + b^2 - y^2)c^2] + \frac{(1+\sigma)(1-2\sigma)}{2(1-\sigma)}\frac{qc-2p_0}{E}c.$$

These three equations represent the sought surface after deformation in the parametrical form with the two variable parameters x, y. By eliminating these two parameters from this system, we shall obtain for the deformed surface under consideration one cartesian equation of an algebraical surface of the fourth order; this equation will contain the coördinate ζ only in the first degree.

15. **Conclusion.** Supposing in all formulas found

$$p_0 = q = 0,$$

we shall obtain the expressions for a particular case of the flexure of a rectangular covering slab with elastically fixed lateral sides, which is only under the influence of the uniformly distributed pressure (p).

This case is analogous to the flexure of a plate given by A. E. H. Love ("A Treatise on the Mathematical Theory of Elasticity", fourth edition, Cambridge, 1927, §307); but A. E. H. Love does not examine the influence of the form, the dimensions, and the conditions of fixing the plate under consideration. Therefore his formulas of unit stresses, supposing $c = h$, coincide in the above particular case with those obtained by the author only for the unit stresses Z_z, X_z, Y_z. The formulas of other unit stresses, as well as the expressions of displacements u, v, w shall be different under the influence of the above factors. The equations of the deformed surface for the median plane shall also be different although they are of the same order, namely of the fourth.

HARBIN POLYTECHNIC INSTITUTE.

SYNTHESIS OF REACTANCE 4-POLES WHICH PRODUCE PRESCRIBED INSERTION LOSS CHARACTERISTICS

Including Special Applications to Filter Design

By S. Darlington*

Outline

Introduction 257
Part I—Open- and Short-Circuit Impedances of Networks with Prescribed Insertion Loss Characteristics 269
Part II—Realization of Open- and Short-Circuit Impedances with Physical Configurations 281
Part III—Special Insertion Loss Functions for Filtering Purposes. 309
Part IV—Dissipative Reactance Networks 335
Illustrative Numerical Design 346

Introduction

Of the various types of electrical networks which are frequently found useful, one of the commonest is the 4-terminal transducer of reactances, more briefly referred to as the reactance 4-pole.[1] In particular, the selective networks or filters which are commonly used for transmitting certain frequencies while blocking others are almost always reactance 4-poles, and these filters form essential parts of most communication systems.

Detailed methods of designing filters and related reactance 4-poles are well known and have been in general use for a considerable period. For the most part these fit into one general filter design scheme which may be referred to as the image parameter theory, since it is based upon

* Bell Telephone Laboratories, Inc. This paper has been accepted as a Doctor's thesis by the Faculty of Pure Science of Columbia University. The manuscript was received by the Editors May 18, 1938.

[1] Throughout this paper, the term 4-pole will be used to indicate a 4-terminal transducer—i.e. a network with two pairs of accessible terminals subject to the restriction that no external connections can be made between terminals of different pairs. The term has been widely used in this sense and also to indicate a network with four terminals to which external apparatus can be connected in any desired manner.

COLE MEMORIAL LIBRARY

the concepts of the image impedances and image transfer constants of 4-terminal networks. It has been found that methods based upon this theory can be used to design practical filters with electrical characteristics meeting any ordinary engineering requirements. In recent years it has become apparent, however, that these filters are sometimes unnecessarily costly. At the same time the question of network cost has become increasingly important.

This paper describes a theory of reactance 4-pole design which differs from the image parameter theory in such a way that it sometimes leads to more advantageous choices of element values in filters of conventional image parameter configurations. While more complicated than the image parameter theory, it sometimes permits the realization of substantially greater network economy or of superior electrical characteristics with no increase in cost. Instead of starting with the concepts of image impedances and image transfer constants, this theory is based upon the general problem in network synthesis of finding reactance 4-poles yielding prescribed insertion loss versus frequency functions when inserted between prescribed resistance terminations. Because of this it has come to be called the insertion loss theory.

While the insertion loss theory applies particularly to filter design, it is capable of more general applications. As a matter of fact, it is theoretically possible to use the theory to design physical reactance 4-poles which, when terminated in prescribed resistances, produce insertion loss characteristics identical with those of any general finite passive 4-poles with the same terminations. Although the possibility of designing at least simple reactance networks on an insertion loss basis is common knowledge, when other than very simple circuits are considered extensive special theory such as that developed in this paper is necessary if hopelessly complex computations are to be avoided.

Although very general applications of the insertion loss method of design are theoretically possible, even with the theory developed here the numerical complications are such as to limit its practical usefulness. In many design problems, for instance, economies in the cost of actually constructing filters might be obtained by using insertion loss designs rather than image parameter designs but these economies frequently are off-set by the added cost involved in obtaining the insertion loss designs. As a result, the insertion loss theory applies principally to the design of filters which are to be made in such large numbers that construction economies will justify high design costs or which must meet requirements not easily satisfied with image parameter designs.

The basic insertion loss theory applies only to networks of pure reactances. Modifications are included, however, permitting the design of certain types of dissipative reactance networks with prescribed loss characteristics. These have greatly increased the practical advantages obtainable by the use of the theory.

Limitations of the Image Parameter Theory

The great simplicity of the image parameter filter theory is obtained by adopting certain definite restrictions which limit to a considerable extent the choice of the element values. It is the possibility of avoiding these restrictions by using the more complicated insertion loss theory that sometimes renders this theory advantageous as a basis of filter design. While the use of the insertion loss theory in filter design normally involves the introduction of other restrictions, these are such as to lead to the optimum choice of element values for meeting certain types of filter specifications. In order to clarify the status of the insertion loss theory, it will be best to introduce at this point a brief description of the fundamentals of the image parameter theory and of the restrictions which it involves.

The image parameter theory is, of course, an outgrowth of the artificial line theories of Pupin (1) and Campbell (2). The image impedances and image transfer constants in terms of which it is developed correspond to the characteristic impedances and propagation constants of these artificial lines. In its most familiar form the image parameter theory deals directly with the so-called composite filters introduced by Zobel (3), which are made up of chains of sections with matched image impedances but different transfer constants. In the more general form developed by Bode (4), however, it deals with the equivalent restrictions upon the image impedances and transfer constants of complete networks, without actually requiring chains of tandem sections.

The image attenuation of a non-dissipative filter is identically zero over finite ranges of frequencies. If the filter is terminated in its image impedances, the corresponding transducer loss will also be zero.[2] Although the image impedances must vary with frequency, they can

[2] By the transducer loss of a filter is meant the difference in level between the received power and the maximum power obtainable from the generator with any passive network. The image attenuation of a filter is equal to its transducer loss when terminated in its image impedances at all frequencies where these impedances are real. The transducer loss of any network terminated in resistances can be obtained from the insertion loss by adding the reflection loss corresponding to the ratio of the terminations.

approximate constant resistances over the greater parts of the ranges of zero attenuation, or theoretical pass bands. Consequently, if these resistances are used as the actual terminations the transducer loss will be small over the ranges of good approximation.

The situation described above leads to the introduction of the following requirement, which is responsible for the fundamental limitations of the image parameter theory: The image attenuation of a non-dissipative filter is required to be identically zero in continuous frequency ranges including those to be freely transmitted between the actual terminations and is required to be other than zero at all other frequencies. For the types of filters commonly encountered these restrictions reduce by almost half the constants which could otherwise be chosen arbitrarily.

Without further investigation, it might appear that good filter characteristics could be obtained only by satisfying these restrictions at least so closely that permissible departures from them would be of no practical interest. Actually, these requirements are not necessary but instead are artificial or arbitrary restrictions leading to a simple design procedure.

As an illustration, suppose that the elements of a filter designed on the image basis are changed from their design values by various amounts of the order of perhaps 10 or 20 percent. In general this will split the theoretical pass band into a number of theoretical pass bands separated by narrow theoretical attenuation bands.[3] The image attenuation corresponding to these theoretical attenuation bands, however, will ordinarily be very moderate. In addition, the actual transducer loss may be only a fraction of the image attenuation. A transducer loss of less than 0.4 decibels may be obtained, for instance, even though the corresponding image attenuation is as high as 2.5 decibels. As a result, modifying the elements may leave unchanged the frequencies included in the *effective* pass band and may result in an actual decrease in the corresponding transducer loss.

When the dissipation required in actual filters has a marked effect the arbitrariness of the restrictions of the image parameter theory becomes much more striking. Under these conditions the image attenuation is generally far from uniform over the frequency ranges in which it is zero on a non-dissipative basis. The special restrictions of the image parameter theory then amount to the requirement of transmission range

[3] In general the theoretical attenuation band will also be split up by narrow theoretical pass bands, but it will be sufficient for purposes of illustration to consider only the splitting of the pass band.

insertion losses approximating characteristics which are anything but ideal.

The above discussion indicates the arbitrariness of the restrictions forming the basis of the image parameter method of filter *design* or *synthesis*. It is well known that the *circuit analysis* problem of determining the operation of a known network can be solved in terms of the image impedances and image transfer constant of the network even though it does not satisfy the special restrictions of the design theory. When these restrictions are abandoned, however, the image parameters are no longer convenient in *design* problems involving the determination of resistance-terminated reactance networks with prescribed properties.

Other Previous Theories

Cauer (5), Gewertz (6) and others have studied the synthesis of perfectly general reactance 4-poles having prescribed open- and short-circuit impedances. These investigations, however, have not yielded useful methods of designing reactance 4-poles with resistance terminations except under the restrictions of the image parameter theory. On the other hand, they have produced such useful information as the necessary and sufficient conditions satisfied by sets of open- and short-circuit impedances corresponding to physical reactance 4-poles. In addition, they have shown that physical networks of certain so-called canonical configurations can readily be designed to have any specified set of impedances satisfying these conditions. The canonical configurations, however, were chosen purely for their general realizability and ease of design and are rarely of practical interest.

The development of the present insertion loss theory started with the previous theory developed by Norton (7) which permits the design of two or more filters producing a constant resistance at their common terminals when connected in series or in parallel at one end. Norton's theory involves the problem of designing a reactance 4-pole terminated in an open- or short-circuit at one end. This amounts to the special case of the insertion loss theory in which one of the specified terminations is zero or infinite. Norton's detailed design procedure, however, assumes a rather restricted type of prescribed insertion loss characteristic which is of little interest except in the design of filters to operate in constant resistance pairs.

Principal Operations Involved in the Design of Filters on an Insertion Loss Basis

Returning now to the subject of this paper, filters can readily be designed on an insertion loss basis provided certain fundamental opera-

tions can be carried out. There is no problem in fitting these operations together into straightforward design procedures but the operations themselves cannot readily be carried out without special mathematical machinery. The principal object of the insertion loss theory is to supply this mathematical machinery.

The logical order to follow in developing the mathematical machinery of the insertion loss theory turns out to be rather different from the order in which the various parts are used in carrying out actual designs. It will therefore be best to motivate the more detailed analysis by introducing at once a brief description of the various fundamental design operations, rather than introducing each operation at the time the corresponding mathematical machinery is taken up.

The distinguishing feature of the insertion loss theory is the use of an exactly prescribed insertion loss versus frequency function to fix the element values of the final network. Since exactly prescribed loss functions are rarely included in design specifications, the first operation in the design of a network on the insertion loss basis is usually the choice of the specific insertion loss function to be obtained. This choice is guided by the following considerations. First, the loss function must be consistent with the design specifications, e.g., it must represent sufficient suppression of the unwanted frequencies. Second, the function must be of a form leading to an economical network as regards the number of elements involved. Third, it must usually correspond to a particular type of network configuration, such as a ladder or a lattice.

After a specific insertion loss function has been chosen, the determination of a corresponding network involves two principal operations. The first of these is the determination of some or all of the open- and short-circuit impedances of the network from the loss function, the second being then the determination of the actual element values from these impedances.[4] It turns out that there are only a finite number of sets of open- and short-circuit impedances corresponding to reactance networks of minimum complexity producing a specific insertion loss function when terminated in a specific pair of resistances. After one of these sets of impedances has been selected, the method of determining the element values of a corresponding network depends upon the type of

[4] This rule, however, is not without its exceptions. In particular, symmetrical lattices with equal terminations are most easily determined from their insertion loss functions by procedures making no direct references to their open- and short-circuit impedances.

configuration chosen, equivalent networks of different configurations frequently requiring quite different methods.[5]

The parasitic dissipation which must be present in actual reactance networks sometimes influences the loss characteristics to a small enough extent to justify neglecting it in the design of the networks. On the other hand, the influence of the necessary dissipation is sometimes so important that it is highly desirable to compensate for it in some manner. The logical way to obtain this compensation is to design dissipative reactance networks which themselves produce more or less exactly prescribed loss functions. Under certain conditions this can be accomplished by modifying the first part of the non-dissipative design procedure in such a way as to obtain the open- and short-circuit impedances corresponding to the removal of the dissipation from the final network.[6] This permits the element values to be computed from the impedances on a strictly non-dissipative basis.

In order to clarify further the general design procedure outlined above, the procedure used in a simple illustrative case will now be described in more detail. It should be borne in mind, however, that the detailed design procedures used in other cases may differ considerably from this illustrative case even though they involve the same general types of operations.

Illustrative Special Case—Choice of Insertion Loss Function

Assuming for the time being that dissipation can be neglected entirely, consider the design of a low-pass filter of ideal reactance elements consistent with the following specifications. First, the insertion loss shall not be greater than a prescribed maximum α_p^0 in an effective pass band extending from zero frequency to a prescribed cut-off frequency f_1^0. Second, the insertion loss shall not be less than a prescribed minimum α_a^0 in an effective attenuation band extending from a second prescribed frequency f_2^0 to infinity. Third, these loss specifications are to be met

[5] As would be expected, when a particular configuration is desired care must usually be exercised to make sure that an appropriate set of impedances is chosen. The various sets corresponding to a single loss function, for instance, usually occur in inverse pairs leading to inverse configurations.

[6] The possibility of doing this depends upon the following situation. When the dissipation required in a network satisfies certain restrictions as regards its variation from element to element the effect of the dissipation on complex impedances and complex voltage ratios corresponds to simple transformations of the functions of frequency representing the impedances and voltage ratios, these transformations being independent of the specific network configuration and element values.

with a network which is electrically symmetrical and is terminated in equal resistances and which has the same configuration as a mid-series low-pass ladder filter of the image parameter theory even though it may have more general element values.[7]

The above specifications on the insertion loss are indicated graphically in Fig. 1A, it being required that the curve of actual insertion loss shall fall within the shaded areas at frequencies less than f_1^0 or greater than f_2^0. The required configuration is indicated in Fig. 2.

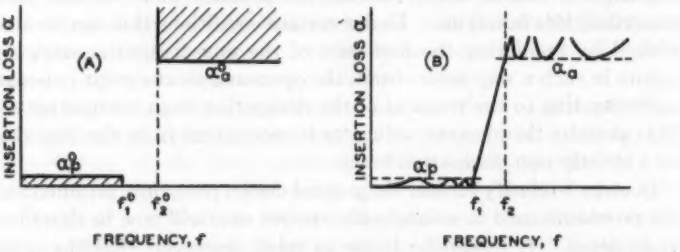


Fig. 1. A) Design specifications—If $f < f_1^0$ or $> f_2^0$, α must lie in a shaded area. B) Form of α characteristic meeting the design specifications.

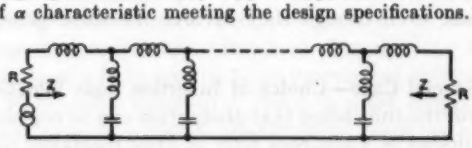


Fig. 2. Configuration to meet specifications of Fig. 1A

The theory to be developed shows that the above specifications on the form of the final network require the insertion loss α to be described by an equation of the form

$$e^{2\alpha} = 1 + \left[S_0 \frac{\omega(\omega_1^2 - \omega^2) \cdots (\omega_\eta^2 - \omega^2)}{(1 - \nu_1^2\omega^2) \cdots (1 - \nu_\eta^2\omega^2)} \right]^2 \tag{1}$$

In this equation, η represents the number of "sections," while S_0 and the ω_σ's and ν_σ's are arbitrary constants. The specific problem is to

[7] It turns out that the requirement that the network shall be either symmetrical or of a type producing inverse image impedances at the two ends tends to lead to efficient filter characteristics. The inverse impedance case is excluded here in order to simplify the illustration. Similarly, the special loss specifications described above are chosen for their simplicity rather than for their generality. Somewhat similar methods can still be used, for instance, when the minimum permissible attenuation band loss varies with frequency.

choose the arbitrary constants in such a way as to satisfy the specifications indicated in Fig. 1A with the smallest possible value of η, i.e., the smallest number of sections.[8]

The best choice of the ω_σ's and ν_σ's turns out to be that leading to the form of loss characteristic indicated in Fig. 1B for the special case of $\eta = 3$. A lengthy analysis will later be outlined which shows that the particular rational "power ratio" function $e^{2\alpha}$ which has the form of (1) and which corresponds to the "equal ripple" type of loss characteristic indicated in Fig. 1B can be obtained by requiring $e^{2\alpha}$ to be determined by the following set of equations relating both $e^{2\alpha}$ and ω to a new variable u.

$$e^{2\alpha} = 1 + (e^{2\alpha_p} - 1)\, sn^2\left[(2\eta + 1)u\frac{K_1}{K}, k_1\right] \qquad (2)^9$$

$$\frac{\omega}{2\pi\sqrt{f_1 f_2}} = \sqrt{k}\, sn(u, k)$$

$$k = \frac{f_1}{f_2}$$

$$q_1 = q^{2\eta+1}$$

In these equations, α_p, f_1, f_2 represent the maximum pass band loss and the limits of the effective pass and attenuation bands, as in Fig. 1B. The symbols K, K_1 and q, q_1 represent constants appearing in the general theory of elliptic functions, K, K_1 being the complete elliptic integrals of the first kind of moduli k, k_1, respectively, and q, q_1 the corresponding elliptic modular constants. These constants are uniquely related to k, k_1 in a manner represented in most elliptic function tables by tabulations of K and $\log_{10} q$ vs $\sin^{-1} k$.

[8] It turns out that choosing the constants in this way will normally lead to a physical network of the form of Fig. 2, although it is possible for difficulties to be encountered if α_p^0 is exceptionally small.

[9] The appearance of elliptic functions in a problem involving an algebraic function producing equal maxima and equal minima is not surprising to those familiar with the elliptic functions appearing in the "Tchebycheff parameter" version of the image parameter theory introduced by Cauer (16). The rational character of $e^{2\alpha}$ considered as a function of ω depends upon the equivalence of $sn\left[(2\eta + 1)u\frac{K_1}{K}, k_1\right]$ to an odd rational function of sn (u, k), which can be compared with the equivalence of sin $[(2\eta + 1)u]$ to an odd polynomial in sin (u). The correspondence of (2) to the particular special case of (1) illustrated by Fig. 1B depends upon the periodic properties of α and ω considered as functions of u.

The insertion loss corresponding to (2) is completely determined by the choice of α_p, f_1, f_2 and the number of sections η, for f_1, f_2 determine k, which determines k_1 through the relation between q and q_1, while k, k_1 determine K, K_1. The corresponding value of the minimum attenuation band loss α_a indicated in Fig. 1B is thus also determined by the choice of α_p, f_1, f_2. If minor approximations are made, this relationship can be expressed in the following form:

$$\alpha_a = [10 \log_{10}(e^{2\alpha_p} - 1) - 10(2\eta + 1) \log_{10}(q) - 12.04] \text{ db} \quad (3)$$

With the help of (3) it is a simple matter to determine the minimum value of η for which the specifications of Fig. 1A can be satisfied, i.e., for which α_p, α_a, f_1 and f_2 can be chosen in such a way that

$$\begin{aligned} \alpha_p &\leqq \alpha_p^0 \qquad f_1 \geqq f_1^0 \\ \alpha_a &\geqq \alpha_a^0 \qquad f_2 \leqq f_2^0 \end{aligned} \quad (4)$$

In this way, a definite set of constants can be chosen which determine, with the help of (2), a loss characteristic leading to a final network of the required type and of the minimum number of sections permitted by the loss specifications.[10]

Illustrative Special Case—Determination of a Network Producing the Insertion Loss Chosen

The open- and short-circuit impedances of reactance networks producing the particular insertion loss function chosen cannot be expressed directly in terms of the loss function itself. They can readily be evaluated, however, in terms of the roots of the corresponding power ratio (1).[11] In the special case under consideration these roots can be computed by means of straightforward formulae derived from the elliptic function relations (2).

If reactance networks of minimum complexity are assumed, and also a definite pair of (equal) terminating resistances, it turns out that there are only four sets of open- and short-circuit impedances consistent with

[10] Any three of the four relations of (4) can be made exact equalities but the fourth will then be an equality only in very special cases. Keeping three of the parameters α_p, α_a, f_1, f_2 fixed and changing the number of sections η makes discrete changes in the fourth parameter. Since some margin is thus usually available even though η has the smallest permissible value, this margin is usually best distributed among the various relations of (4).

[11] If the requirement of a loss function appropriate for a symmetrical network is abandoned it is usually necessary to include also the roots of a linear function of the power ratio $e^{2\alpha}$.

(1). Of these only one set is appropriate for the mid-series low pass ladder configuration of Fig. 2. Formulae for computing the element values of this configuration from the impedances will be included in the more detailed development of the insertion loss theory.[12]

Illustrative Special Case—Compensation for Effects of Dissipation

Turning now to the question of compensating for effects of dissipation, consider first the problem of attempting to design a network of reactance elements of prescribed dissipativeness producing exactly the same insertion loss and phase as the non-dissipative filter considered previously. Assuming such a network to be possible, consider the insertion loss and phase that would be produced by the pure reactance network formed by simply removing the dissipation from all the elements. Provided the elements are all to be equally dissipative, it turns out that this new loss and phase can be evaluated directly from the original loss and phase and the extent of the dissipation, with no further information as to the actual element values or specific configuration.[13] Thus any pure reactance network which could be designed to produce this "predistorted" loss and phase would produce the original loss and phase upon the addition of the required dissipation.

As described above, the "predistorted" insertion loss and phase cannot actually be produced by a pure reactance network. On the other hand, if the required dissipation varies properly with frequency and is not too great, the predistortion can be modified in such a way as to lead to a physical network of the configuration of Fig. 2 which will produce a close approximation to the sum of the original insertion loss and an added constant loss when the required dissipation is added. This modified predistortion amounts to nothing more than the addition of a

[12] When the insertion loss function is not necessarily appropriate for a symmetrical network there are usually more than four sets of open- and short-circuit impedances corresponding to networks of minimum complexity. There may be a considerable variety, for instance, realizable with various networks having a single configuration, such as the configuration of Fig. 2.

In the special case of symmetry under consideration the element values of the final ladder may be computed directly from the equivalent lattice rather than from open or short-circuit impedances. The formulae for doing this, however, are derived by expressing the open-circuit impedances in terms of the impedances of the lattice arms.

[13] Actually the corresponding "complex insertion voltage ratio" is more convenient to use than the insertion loss and phase. The original voltage ratio can be readily evaluated in terms of the roots of (1) used in computing the open- and short-circuit impedances on a non-dissipative basis.

constant to each root of the complex voltage ratio function corresponding to the original loss and phase plus the addition of a constant factor representing an added constant loss.[14] Although the general configuration of Fig. 2 is obtained, however, the network yielded by the predistortion method will not satisfy the previous requirement of electrical symmetry nor will the terminations be equal.[15]

Major Divisions of the Theoretical Development

As was indicated previously, the logical order to follow in developing the insertion loss theory is rather different from the order in which the various parts are used in carrying out actual designs. Specifically, the logical procedure is to divide the development of the theory into the sequence of four parts described briefly below, which represent rather different theoretical problems even though they may all be employed in designing a single filter.

Part I deals with the determination of the open- and short-circuit impedances of reactance 4-poles corresponding to prescribed insertion loss characteristics and pairs of terminations. The necessary and sufficient conditions upon insertion loss characteristics yielding physical sets of impedances are established as well as the procedure for determining the impedances.

Part II deals with the determination of reactance 4-poles of special configurations from prescribed open- and short-circuit impedances. These include a new canonical reactance 4-pole made up of tandem sections and also networks which have the same configurations as the familiar lattices and ladders of the image parameter theory but which are not required to satisfy the restrictions on actual element values imposed by the image parameter theory.

Part III deals with the problem of choosing specific insertion loss functions for filtering purposes. In particular, it deals with certain special types of loss functions which satisfy the conditions for physical realizability established in Part I, which represent the optimum choice

[14] The roots of the voltage ratio are determined in evaluating it from the roots of (1).

[15] When the dissipation varies from element to element or varies improperly with frequency, partial compensation can be obtained by assuming dissipation of the proper type representing an average of the actual dissipation. It is also possible to reduce the restrictions on the dissipation assumed if the non-dissipative design procedure is further modified. It is possible for instance, to assume all coils to be equally dissipative and all condensers to be equally dissipative to a different extent.

of arbitrary constants for meeting certain types of filter requirements, and which lead to simplifications in the more general design procedure.

Part IV deals with modifications of the previous theory permitting the design of certain types of dissipative networks with prescribed insertion loss characteristics. It is shown that within certain severe limits it is possible to design dissipative filters with loss characteristics differing appreciably from those of non-dissipative filters of the same configuration only by added constant losses.

Because of the extent of the material to be covered, detailed derivations and proofs will be only outlined in developing the four parts of the theory described above. In addition, a variety of useful but unessential modifications of the theory will be omitted entirely. These modifications have to do with the application of the theory to various special design problems rather than forming a part of the central theory itself.

Fig. 3. Reactance 4-pole with resistance terminations

Part I. Open- and Short-Circuit Impedances of Networks with Prescribed Insertion Loss Characteristics

Statement of Problem

As is indicated in the above outline, the first two major divisions of the insertion loss theory are not restricted to the design of networks specifically for filtering purposes. Finite 4-terminal networks are assumed and these are required to be made up of pure reactances and to be terminated in constant resistances as in Fig. 3, but they are not necessarily required to have filter-like characteristics. The terminating resistances R_1, R_2 are assumed to be prescribed in each case and also the exact function of frequency representing the insertion loss α due to the presence of the reactance network. The specific problem considered is the design of the reactance network in accordance with these specifications and subject to the requirement of physical realizability.

The first division of the theory solves the problem of determining sets of open- and short-circuit impedances describing physical reactance

networks which would produce the prescribed insertion loss when inserted between the prescribed terminations. When combined with Cauer's method (5) of realizing prescribed impedances with reactance 4-poles of his so-called canonical configurations, it amounts to an academic solution of the general design problem described above. The design of more important equivalent configurations will be described in the second division of the theory.

To insure the possibility of physical corresponding networks the prescribed insertion loss function α must be assumed to satisfy certain necessary and sufficient conditions. By definition, α satisfies the equation

$$e^{2\alpha} = \left| \frac{V_{20}}{V_2} \right|^2 \tag{5}$$

in which V_{20} and V_2 represent the complex voltages received before and after the insertion of the network between the prescribed terminations. Since it represents the square of the magnitude of the "complex insertion voltage ratio" $\frac{V_{20}}{V_2}$, the "insertion power ratio" $e^{2\alpha}$ must be an even rational function of frequency with real coefficients and must be positive at all real frequencies. Because the power delivered by a generator of voltage e and internal resistance R_1 cannot exceed $\frac{e^2}{4R_1}$, while the received power corresponding to V_{20} is $\frac{R_2 e^2}{(R_1 + R_2)^2}$, the power ratio $e^{2\alpha}$ must not only be positive at real frequencies but must not be less than $\frac{4R_1R_2}{(R_1 + R_2)^2}$. In other words, the relation

$$e^{2\alpha} \geqq \frac{4R_1R_2}{(R_1 + R_2)^2} \tag{6}$$

must be satisfied at all real frequencies.

The above conditions upon $e^{2\alpha}$ are obviously necessary whether or not the network inserted between the terminations must be made up entirely of reactances. It will be shown, however, that these conditions are sufficient to insure the existence of physical corresponding networks of the pure reactance type.

Form of Solution

It turns out that the open- and short-circuit impedances of a reactance network depend upon the insertion loss in a manner involving a rather

complicated combination of the roots of the insertion power ratio and a related function, which are normally of fairly high degree. As a result, explicit equations expressing the impedances directly in terms of the insertion loss cannot well be obtained. On the other hand, a simple sequence of formulae has been derived which indicates the relationship and which permits the impedances to be calculated in numerical cases.

Instead of actually deriving the formulae, it will be sufficient to begin by merely stating them without proof. A derivation is unnecessary in that elementary circuit analysis can be used to show that any networks producing the impedances determined by these formulae would yield the prescribed insertion loss when inserted between the prescribed terminations. After the statement of the formulae, however, it will be necessary to demonstrate that the impedances can actually be realized physically.

Determination of Polynomials N and P from $e^{2\alpha}$

To determine the open- and short-circuit impedances corresponding to the prescribed insertion loss α, the power ratio $e^{2\alpha}$ is first expressed in the form

$$e^{2\alpha} = \frac{N}{P^2} \tag{7}$$

in which N and P denote even polynomials in the familiar variable p representing $2\pi if$ and are required to have real coefficients. An expression of this type can be readily obtained, since $e^{2\alpha}$ must be an even rational function of the frequency f, while f^2 is proportional to $-p^2$. If the simplest rational fraction expression for $e^{2\alpha}$ does not have the form $\frac{N}{P^2}$ it is only necessary to multiply numerator and denominator by identical factors. As a matter of fact, N and P can always be chosen in a variety of ways, because of the possibility of associating the squares of arbitrary identical factors with N and P^2.

Determination of Polynomials A and B from N

After N and P have been determined, a second pair of polynomials A and B must be evaluated. These are even polynomials with real coefficients defined by the pair of equations

$$\begin{aligned} N &= J^2(p_1^2 - p^2)(p_2^2 - p^2) \cdots (p_n^2 - p^2) \\ A + pB &= J\,(p_1 - p\,)(p_2 - p\,) \cdots (p_n - p) \end{aligned} \tag{8}$$

in which J must be a real constant, all the p_σ's must have zero or negative real parts, and any complex p_σ's must occur in conjugate pairs. The fact that N must be positive at real frequencies in order for the power ratio $\frac{N}{P^2}$ to be positive is sufficient to insure the possibility of computing J and the p_σ's from their squares, as determined by N, in such a way as to meet the special restrictions imposed upon them.[16] As a matter of fact, there are always two solutions, although only two, because of the arbitrariness of the sign of J. The individual polynomials A and B can obviously be determined by expanding $(A + pB)$ and associating even powers of p with A and odd powers with pB.

The following expression for the complex insertion voltage ratio offers an alternative definition of A and B and indicates their physical significance:[17]

$$\frac{V_{20}}{V_2} = \frac{A + pB}{P} \tag{9}$$

That this voltage ratio corresponds to the original power ratio $e^{2\alpha}$ can be demonstrated by noting that the even polynomials A, B, P are real at real frequencies, while p is imaginary. This requires the power ratio to be

$$e^{2\alpha} = \frac{A^2 - p^2B^2}{P^2} \tag{10}$$

[16] Since the p_σ's are determined by extracting the square roots of the p_σ^2's determined by N, the signs can be so chosen as to meet the real part requirement. Complex p_σ's with finite real parts will then automatically occur in conjugate pairs corresponding to conjugate p_σ^2's. Pure imaginary p_σ's, which have zero real parts, can also be chosen in conjugate pairs provided the corresponding negative real p_σ^2's occur in identical pairs. Single negative real p_σ^2's cannot be encountered since they represent real frequencies at which N and therefore $\frac{N}{P^2}$ change sign.

[17] In special cases, it may be possible to cancel identical factors out of the numerator and denominator of this expression. Unless the factors are constants or have roots only at real frequencies, however, their cancellation destroys the evenness of the denominator. The even and odd parts of the numerator then lose their utility in regard to the computation of the open- and short-circuit impedances.

The determination of a voltage ratio from a power ratio by (8) and (9) was introduced by Norton (7) in connection with his theory of filters terminated at one end in open- or short-circuits, which forms a part of his theory of constant-resistance groups of filters.

Since reversing the sign of p does not change the even polynomials A and B, the second equation of (8) requires

$$A - pB = J(p_1 + p)(p_2 + p) \cdots (p_n + p) \tag{11}$$

Hence forming the product of $(A + pB)$ and $(A - pB)$ indicates

$$A^2 - p^2B^2 = N \tag{12}$$

and reduces (10) to (7).

The voltage ratio (9) would correspond to the original power ratio even though roots of $(A + pB)$ had positive real parts, provided complex roots occurred in conjugate pairs. The exclusion of roots with positive real parts is merely a condition necessary for physical realizability. This condition is necessary in that the roots represent modes of free oscillation of the complete network consisting of the reactance 4-pole and its terminations.

Determination of Polynomials A' and B' from N and P

In addition to A and B, a very similar pair of polynomials A' and B' are needed. They are even polynomials with real coefficients defined by the following pair of equations, comparable to (8):

$$N - \frac{4R_1R_2}{(R_1 + R_2)^2} P^2 = J'^2(p_1'^2 - p^2)(p_2'^2 - p^2) \cdots (p_n'^2 - p^2)$$

$$A' + pB' = J'(p_1' - p)(p_2' - p) \cdots (p_n' - p) \tag{13}$$

In this case, J' must be real and any complex p_σ''s must occur in conjugate pairs but any of the p_σ''s are permitted to have positive real parts. These conditions can always be satisfied because of the fact that (6) and (7) require $\left(N - \frac{4R_1R_2}{(R_1 + R_2)^2} P^2\right)$ to be non-negative at real frequencies. In general, there are actually a number of solutions for A' and B' because of the arbitrariness of the signs of the p_σ''s as well as that of J'.

In the case of A' and B' there is again an alternative definition indicating physical significance, corresponding to equation (9) defining A and B. Suppose Z_1 and Z_2 represent the input impedances of the terminated network measured at the ends terminated in R_1 and R_2, respectively, as indicated in Fig. 4. Then A' and B' can be defined by the equations[18]

$$\frac{R_1 - Z_1}{R_1 + Z_1} = \frac{A' + pB'}{A + pB} \qquad \frac{R_2 - Z_2}{R_2 + Z_2} = \frac{-A' + pB'}{A + pB} \tag{14}$$

[18] In special cases, it may be possible to cancel out common factors from

The square of the magnitude of $(\pm A' + pB')$ is $(A'^2 - p^2B'^2)$, which in turn satisfies the equation

$$A'^2 - p^2B'^2 = N - \frac{4R_1R_2}{(R_1 + R_2)^2} P^2 \tag{15}$$

corresponding to (12). Hence (14) requires

$$\left|\frac{R_1 - Z_1}{R_1 + Z_1}\right|^2 = \left|\frac{R_2 - Z_2}{R_2 + Z_2}\right|^2 = 1 - \frac{4R_1R_2}{(R_1 + R_2)^2} e^{-2\alpha} \tag{16}$$

This expression can be checked by examining the relation between the input power and input impedance and recalling that input and output powers are identical in the case of a reactance network.

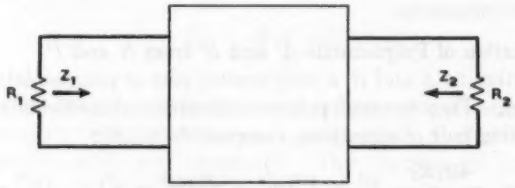


Fig. 4. Input impedances Z_1 and Z_2

The possibility of roots of $(A' + pB')$ with positive real parts is due to the fact that they are the roots of the non-physical impedance $(R_1 - Z_1)$. The simple relation between the two reflection coefficients indicated in (14) can be checked by simple circuit analysis. The following important relation between the A's and B's represents a combination of (12) and (15):

$$A'^2 - p^2B'^2 = A^2 - p^2B^2 - \frac{4R_1R_2}{(R_1 + R_2)^2} P^2 \tag{17}$$

Computation of the Open- and Short-Circuit Impedances from the A's, B's and P

In terms of the A's, B's and P the desired open- and short-circuit impedances are determined by the following formulae:

---

one or both of the equations for the reflection coefficients as in the case of the voltage ratio (9). If the denominators are not $(A + pB)$, however, the numerators lose their direct application to the computation of the open- and short-circuit impedances.

$$Z_{01} = R_1 \frac{A - A'}{p(B + B')} \qquad Z_{S1} = R_1 \frac{p(B - B')}{A + A'}$$

$$Z_{02} = R_2 \frac{A + A'}{p(B + B')} \qquad Z_{S2} = R_2 \frac{p(B - B')}{A - A'} \tag{18}$$

$$Z_{012} = \frac{-2R_1R_2}{(R_1 + R_2)} \frac{P}{p(B + B')} \qquad Z_{S12} = \frac{(R_1 + R_2)}{2} \frac{p(B - B')}{P}$$

In these formulae Z_{01}, Z_{S1}, Z_{02}, Z_{S2} are the familiar open- and short-circuit driving-point impedances of the reactance network as measured at the terminals next to terminations R_1 and R_2, respectively. Z_{012} and Z_{S12} are the less familiar open- and short-circuit transfer impedances defined in Fig. 5.[19] That the impedances determined by (18) correspond to the original power ratio and to the related voltage ratio and reflection coefficients of (9) and (14) can be checked by means of simple circuit analysis.[20]

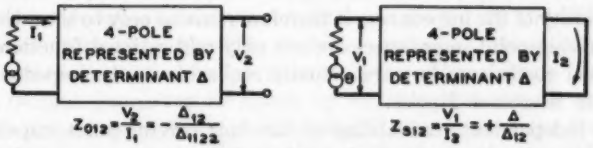


Fig. 5. Open- and short-circuit transfer impedances Z_{012} and Z_{S12}

Physical Realizability of the Impedances Determined by (18)

The necessary and sufficient conditions for the realizability of a set of open- and short-circuit impedances with a physical reactance 4-pole can be stated as follows: First, the four driving-point impedances Z_{01}, Z_{02}, Z_{S1}, Z_{S2} must be separately realizable as the impedances of reactance 2-poles. Second, the two transfer impedances Z_{012}, Z_{S12} must be odd rational functions of p with real coefficients even though they need not be realizable as 2-terminal impedances. Finally, the various

[19] In terms of more conventional notation, these impedances would be referred to as the transfer impedance and the reciprocal of the transfer admittance. The more specific description is called for in the design of tandem section networks such as those to be described in Part II, which sometimes involves the use of short-circuit impedances or open-circuit admittances.

[20] The demonstration may also require the use of equation (17) relating the A's and B's if certain possible expressions for the voltage ratio and reflection coefficients in terms of the impedances are chosen as a starting point.

impedances must be inter-related by the following identities:

$$Z_{01}Z_{02} - Z_{012}^2 = Z_{01}Z_{S2} = Z_{02}Z_{S1} = -Z_{012}Z_{S12} \tag{19}$$

The above conditions differ from the more familiar necessary and sufficient conditions introduced by Cauer (5) in that the requirement of 2-pole realizability is imposed explicitly on all four driving-point impedances, instead of on only two, while Cauer's requirement on the residues of a transfer impedance is abandoned. It turns out that the requirement of four physical 2-pole impedances, together with (19), is sufficient to insure the satisfaction of Cauer's residue requirement. This can be demonstrated by analyzing the behavior of (19) in the neighborhood of a pole of Z_{012}.

An inspection of (18) shows that the expressions for the driving-point and transfer impedances are all odd rational functions of p with real coefficients. The required identities (19) are readily shown to be satisfied by merely replacing the impedances by the corresponding expressions of (18) and then making use of (17). To prove the complete realizability of the impedances it therefore remains only to show that the four driving-point impedances are not only odd rational functions of p with real coefficients but are actually realizable as the impedances of separate reactance 2-poles.

The independent realizability of the four driving-point impedances can be demonstrated with the help of the following theorem. If A_x and B_x are even polynomials in p with real coefficients such that $(A_x + pB_x)$ has no roots with positive real parts, then $\dfrac{pB_x}{A_x}$ is realizable as the impedance of a physical reactance 2-pole. The truth of this theorem is indicated by the argument outlined below in terms of the theory of positive real functions, or impedances of general 2-poles, as developed by Brune (8).

In the first place, the combination $\dfrac{A_x}{A_x + pB_x}$ must be a positive real function because of the fact that its poles lie in the left half of the p plane while its resistance part, $\dfrac{A_x^2}{A_x^2 - p^2B_x^2}$, is non-negative on the real frequency axis.[21] Hence the reciprocal, $\left(1 + \dfrac{pB_x}{A_x}\right)$, is a positive real

[21] The polynomial $A_x + pB_x$ is permitted to have roots on the real frequency axis as well as in the left half of the p plane but these are always roots of A_x and pB_x and produce no pole of $\dfrac{A_x}{A_x + pB_x}$.

function. This requires $\frac{pB_z}{A_z}$ to be a positive real function, since subtracting a real constant from a positive real function leaves a positive real function provided the real part is still non-negative on the real frequency axis. Finally, the fact that the positive real function $\frac{pB_z}{A_z}$ is odd requires it to be realizable as the impedance of a 2-pole of the reactance type.

Dividing the four driving-point impedances of (18) by R_1 or R_2 and then replacing the open-circuit functions by their reciprocals yields the four functions corresponding to different choices of the signs in $\frac{p(B \pm B')}{(A \pm A')}$. Thus the four driving-point impedances will be realizable with separate reactance 2-poles provided $[(A \pm A') + p(B \pm B')]$ has no roots with positive real parts. That this condition is satisfied can be demonstrated by showing $\frac{[(A \pm A') + p(B \pm B')]}{A + pB}$ to be a positive real function. The positive realness of this new function is proved by the following conditions: First, the poles all lie in the left half plane, since $(A + pB)$ has no roots with positive real parts; second, the real part is non-negative at real frequencies, as can be shown by recalling that the A's and B's are real at real frequencies, by evaluating the real part of the function on this basis, and by then using equation (17) relating the A's and B's.[22]

Multiplicity of Solutions

As was indicated previously, a multiplicity of solutions for the impedances can be obtained in a variety of ways. In the first place, squares of identical factors can always be associated with the polynomials N and P^2, in the power ratio $\frac{N}{P^2}$. These factors cancel out in the final impedance formulae only if they are constants or if all their roots occur at real frequencies. In addition, the dependence of A, B, A', and B' upon N and P leaves the sign of $(A + pB)$ arbitrary and normally permits a number of solutions for A' and B'.

The choice of N and P determines the complex insertion voltage ratio $\frac{A + pB}{P}$ except as to sign, while the voltage ratio in turn determines the complexity of corresponding networks, which is normally the same

[22] Roots of $(A + pB)$ on the real frequency axis turn out to be roots of A, pB, A', and pB' and produce no poles of $\frac{[(A \pm A') + p(B \pm B')]}{A + pB}$.

for all choices of A' and B' permitted by N and P.[23] Making the degree of N as low as is permitted by the prescribed insertion loss naturally makes the degree of the voltage ratio as low as possible and leads to the simplest networks. When the squares of identical factors with roots at complex frequencies are combined with N and P^2, more complicated networks producing a modified voltage ratio are obtained. Some of these networks correspond to the addition of familiar all-pass phase-shifting sections to the networks of minimum complexity. An additional multiplicity of networks is normally possible, however, corresponding to impedances which cannot be realized in this way.

The analysis outlined above indicates the existence of a number of physical solutions for the impedances, all assuming a particular form. On the other hand, it does not show that there may not be other solutions of quite different forms. A more complicated analysis, however, shows that there can be no such alternative solutions. In other words, it can be demonstrated that the proper choice of N and P and of the corresponding A's and B's permits the determination of any set of impedances realizable with a reactance network corresponding to the prescribed insertion loss function. This will not be demonstrated here, for the object is merely to show that at least some physical sets of impedances can be obtained.

Practical Design Procedure

For practical purposes it is usually desirable to restrict the insertion power ratio $e^{2\alpha}$ somewhat further than is necessary merely to insure physical realizability. Suppose that identical factors other than constants or the simple factor p^2 must be combined with numerator and denominator of the simplest expression for $e^{2\alpha}$ in order to obtain the form $\frac{N}{P^2}$. It turns out that the corresponding network complexity will then be the same as though these added factors were not identical.[24] On the other hand, requiring the factors to be identical usually leads to less desirable loss characteristics than could be obtained by making

[23] In very special cases, special choices of A' and B' may lead to networks of complexities which are less than normal for the degree of the voltage ratio.

[24] The reason for this is as follows: If Δ is the determinant of the complete network formed by the reactance 4-pole together with its terminations, the insertion power ratio is proportional to $\frac{|\Delta|^2}{|\Delta_{12}|^2}$, in which the cofactor Δ_{12} is independent of the terminations and therefore is an odd or even polynomial in p. Thus the denominator of the simplest rational fraction expression for $e^{2\alpha}$ is the square of an even or odd polynomial in the frequency except when the element values are such that Δ and Δ_{12} have roots in common.

them different. Thus it is generally desirable to require $e^{2\alpha}$ to be so chosen that no identical factors other than constants or the simple factor p^2 need be combined with the simplest corresponding rational fraction in forming $\frac{N}{P^2}$. This merely requires all the poles of $e^{2\alpha}$ to occur in identical pairs.

After the power ratio has been chosen the corresponding open- and short-circuit impedances can be computed in a straightforward manner by means of the formulae described above. Corresponding networks of the canonical type devised by Cauer (5) can then readily be designed from the partial fraction expansions of various of these impedances or of the corresponding admittances. More frequently, however, it is preferable to use equivalent tandem section networks, which can be designed by methods to be described in the next section.

In the determination of the impedances considerable labor may be encountered in the extraction of the roots of the polynomials N and $\left(N - \frac{4R_1R_2}{(R_1+R_2)^2}P^2\right)$, which are needed in the formation of the polynomials $(A + pB)$ and $(A' + pB')$ in accordance with (8) and (13). The required root extraction can now be expedited, however, by the use of machines which have recently been constructed for the determination of the roots of any polynomials of reasonable degree.[25] In the design of networks specifically for filtering purposes, moreover, the special polynomials encountered are usually such that the greater part of the root extraction labor can be avoided by the use of special methods to be described in Part III.

The theory described above assumes non-dissipative reactance networks. The distortion due to the parasitic dissipation which must be present in actual networks can be readily estimated as soon as the complex voltage ratio is determined provided the dissipation is uniformly distributed. This can be accomplished by making use of the theory of uniformly dissipative networks as developed by Mayer (10) and Bode, which shows how the effects of uniform dissipation can be estimated from the derivative of the insertion phase.[26] Ways of avoiding distortion due to dissipation will be described in Part IV, which will be devoted to methods of designing certain types of dissipative networks producing prescribed insertion loss characteristics.

[25] See, for instance, the description of Fry's isograph recently published by Dietzold (9).

[26] Bode developed a theory similar to Mayer's independently but published only a brief reference to it. This reference is in the paper "Ideal Wave Filters" (11) which he published in collaboration with Dietzold.

One Open- or Short-Circuit Termination

Reactance networks terminated at one end in open- or short-circuits can be considered as limiting cases of the previous networks in which R_1 or R_2 approaches zero or infinity. They are of interest for two principal reasons. For one thing, it is sometimes advantageous to approximate actual open-circuit terminations by means of vacuum tubes. In addition, when groups of filters are to be combined in series or in parallel at one end, it is sometimes best to design the separate filters as though they were to be terminated in open- or short-circuits at their common terminals.[27]

Although the general methods described above can still be used, difficulties with the specific impedance formulae are experienced when one termination approaches zero or infinity. It turns out that all but one of the possible solutions for the polynomials A' and B' permitted by (13) lead to element values which approach zero or infinity with the termination. Even with the one permissible choice most of the open- and short-circuit impedance formulae become indeterminate.[28]

More careful analysis yields the following special relations.[29] If termination R_2 is infinite

$$Z_{01} = R_1 \frac{A}{pB} \qquad Z_{012} = -R_1 \frac{P}{pB} \tag{20}$$

$$R_{Z2} = R_1 e^{-2\alpha}$$

in which R_{Z2} is the resistance measured at the open-circuited end of the terminated network. If termination R_2 is zero

$$Z_{S1} = R_1 \frac{pB}{A} \qquad Z_{S12} = R_1 \frac{pB}{P} \tag{21}$$

$$G_{Y2} = \frac{1}{R_1} e^{-2\alpha}$$

[27] This is the method introduced by Norton (7) for the design of constant-resistance groups of filter. It can also be used to design more efficient filter groups producing approximately constant resistances or conductances at their common terminals at pass band frequencies, the susceptances or reactances being later corrected by the addition of reactance 2-poles.

[28] Some such indeterminacy is necessary, since any arbitrary impedance can be connected in parallel with a short-circuit termination or in series with an open-circuit termination without affecting the operation of the circuit.

[29] The formulae for Z_{01}, Z_{S1}, R_{Z1}, G_{Y2} are exhibited by Norton (7) as a part of his theory of constant resistance groups of filters.

in which G_{Y2} is the conductance measured at the short-circuited end of the terminated network. Z_{012}, Z_{S12} are the open- and short-circuit transfer impedances while Z_{01}, Z_{S1} are the open- and short-circuit driving-point impedances corresponding to the end of the network terminated in the finite resistance R_1. The polynomials A, B and P are determined from the power ratio exactly as though termination R_2 were finite.

If minimum complexities are assumed, networks can be designed with no additional impedance data. This can be shown by ananalysis of the relations between the open- and short-circuit impedances of physical reactance 4-poles. It will also be clarified by the ladder network theory which will be considered in the next section.

Part II. Realization of Open- and Short-Circuit Impedances with Physical Configurations

Methods have been developed for designing reactance 4-poles of a variety of different configurations producing prescribed open- and short-circuit impedances, such as impedances determined from insertion loss functions by the method described in Part I. Some of these configurations are said to be canonical in that they can be designed to have any set of impedances realizable with physical reactance networks. Others are less general but can frequently be used more advantageously or else amount to special cases of the general configurations for which simpler design methods have been developed.

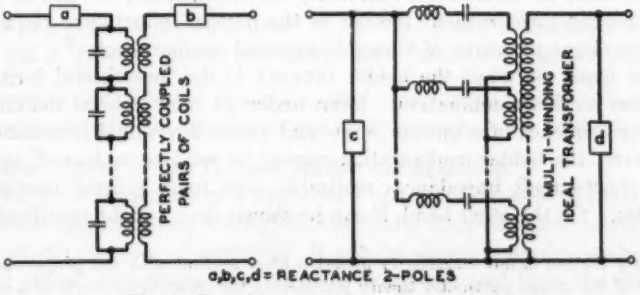


Fig. 6. Cauer's canonical reactance 4-poles

Cauer's Canonical Configurations

The simplest general design theory applies to the design of the two canonical reactance 4-poles introduced by Cauer (5). These have the

configurations indicated in Fig. 6. They are designed in terms of partial fraction expansions of the three open-circuit impedances or short-circuit admittances by noting certain relations between partial fractions of the different expansions and the correspondence of these partial fractions to the various network branches. Since Cauer has described these networks very completely, it will not be necessary to describe them in more detail here.

A Canonical Tandem Section Configuration

Cauer's canonical configurations are of particular interest only in theoretical studies of the properties of reactance 4-poles. When it comes to actual construction it is almost always preferable to use equivalent circuits consisting of simple networks or sections connected in the tandem manner indicated in Fig. 7.[30] In the case of selective networks or filters the use of tandem sections is usually a practical necessity. In the first place, the use of tandem sections permits reasonable approximations to theoretical transmission characteristics to be obtained with much less precise adjustment of the elements to their theoretical values.

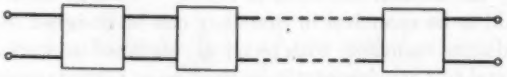


Fig. 7. Tandem 4-poles

In addition, the filters most commonly encountered can be built in tandem section form without the use of the mutual inductances required in equivalent networks of Cauer's canonical configurations.[31]

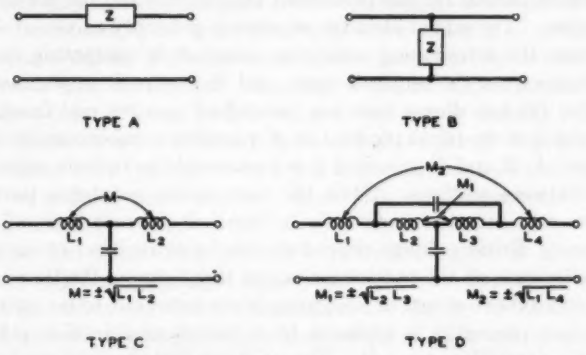
For many purposes the ladder network is the most useful form of tandem section combination. Even under its most general definition as any sequence of alternate series and shunt 2-terminal impedances, however, the ladder configuration cannot be used to realize all open- and short-circuit impedances realizable with more general reactance 4-poles. On the other hand, it can be shown that a slight modification

[30] By section is not necessarily meant a filter section with the properties of those of the image parameter theory but merely the constituent parts of a combination of 4-poles when these are to be connected in tandem.

[31] The requirement of mutual inductances adds substantially to the difficulty of building a network. If perfect coupling between coils is called for, it can be only approximated. If less than perfect coupling is required, it is difficult to obtain simultaneous adjustments of the self and mutual inductances.

or generalization of the ladder configuration can always be realized and hence can be referred to as a canonical reactance 4-pole. This configuration is defined as any tandem combination of sections included in the four types A, B, C, D indicated in Fig. 8, plus possibly an ideal transformer in tandem at one end of the network.[32]

In many cases sections of all four types are not required in the canonical network described above and frequently the ideal transformer does not appear. It also turns out that series inductances can often be included in the series branches, or networks of type A, in such a way that the inductive coupling in sections of types C and D can be less than perfect or can even be completely replaced by separate self-inductances in the shunt branches of these sections. In most filters, for instance,



Z = a general reactance 2-pole. Types C and D may be dissymetrical.

Fig. 8. Types of tandem sections yielding a canonical reactance 4-pole

neither sections of type D nor the ideal transformer appear, while the inductive coupling in sections of type C can almost always be eliminated.

General Theory of the Tandem Section Configuration

The input impedance of a reactance 4-pole terminated at one end in a constant resistance can readily be determined from the open- and short-circuit impedances. Conversely, it can be shown that the open- and short-circuit impedances can normally be determined from the impedance function corresponding to a prescribed terminating resistance,

[32] Other canonical networks of similar tandem sections can be obtained by modifying somewhat sections of types C and D in accordance with the principles of inverse networks and frequency transformations. It will be sufficient to consider here, however, only the particular type of tandem combination described above.

except for an obvious ambiguity as to the signs of the transfer impedances corresponding to the possibility of interchanging the input or output terminals. The only exceptions to the rule correspond to values of the termination leading to impedances of reduced degree, and there can be only a finite number of such values for any one network.[33]

In accordance with the above principle, one way to obtain prescribed open- and short-circuit impedances of the reactance type is to design a resistance-terminated reactance 4-pole as a 2-pole producing an input impedance computed from the open- and short-circuit impedances. The addition of the proper ideal transformer to the terminated end of the 4-pole will yield the prescribed open- and short-circuited impedances provided the design is so carried out that the degree of the input impedance is normal for the prescribed impedances and for the final configuration. The added ideal transformer is generally necessary in order to obtain the terminating resistance assumed in computing the input impedance from the required open- and short-circuit impedances.

Brune (8) has shown how any prescribed positive real function can be realized as the input impedance of a tandem combination of sections of types A, B and C provided it is permissible to include series resistances between sections. When the roots of the resistance part of the positive real function occur only in identical pairs and are all real or imaginary, Brune's design procedure can be carried out in such a way as to eliminate all the resistances except that forming the termination.[34] The requirement of real or imaginary roots turns out to be unnecessary if Brune's procedure is replaced by a recent modification permitting sections of type D to be used. The configuration obtained can be shown to be such that impedance functions of the same degree as the prescribed function would be obtained with general values of the terminating resistance, i.e., the prescribed function does not represent a special case

[33] In terms of the determinant Δ of the reactance 4-pole and various of its cofactors, the input impedance is $\dfrac{\Delta + R_2\Delta_{22}}{\Delta_{11} + R_2\Delta_{1122}}$. Except in special cases in which the cancellation of identical factors leads to reductions in the degree of the impedance function, this requires $(\Delta + R_2\Delta_{22})$ and $(\Delta_{11} + R_2\Delta_{1122})$ to be proportional to $Kp^n(A_1 + pB_1)$ and $Kp^n(A_2 + pB_2)$, respectively, in which $(A_1 + pB_1)$ and $(A_2 + pB_2)$ are the numerator and denominator of the simplest rational fraction expression for the impedance function. The necessary evenness and oddness of the determinants permits of only two corresponding solutions for the open- and short-circuit impedances. Since it turns out that passing from one solution to the other reverses the sign of Δ_{12}', only one solution can be physical.

[34] The resistance part of the function is defined as the even part and represents the real part only at real frequencies.

of reduced degree.[35] Thus a tandem section equivalent can readily be found corresponding to any reactance 4-pole which can produce an input impedance such that the degree is normal for the configuration and such that all the roots of the resistance part occur in identical pairs. This turns out to be a property of all physical reactance 4-poles.[36]

The analysis outlined above proves the canonical nature of the reactance 4-pole formed by adding an ideal transformer to one end of a tandem combination of sections of types *A*, *B*, *C*, *D*. It can also be shown that the two-terminal impedance formed by closing one pair of terminals through a resistance termination constitutes a canonical general 2-pole. It was indicated above that this type of 2-pole can correspond to any positive real function provided the roots of the resistance part occur only in identical pairs. The restriction on the positive real function turns out to be unnecessary, however, when it is permissible to use special values of the terminating resistance leading to impedance functions of reduced degree.[37] Because of the canonical nature of the reactance 4-pole, this can be demonstrated by merely showing how it is possible to find a set of physical open- and short-circuit impedances of the reactance type leading to an input impedance represented by any prescribed positive real function. How this can be accomplished is indicated below.

Suppose that a reactance network is to be so designed as to produce a prescribed input impedance at one end when terminated at the far end in a prescribed resistance. Because of the identity of the input and received powers, the insertion loss that would be obtained upon terminating the input end in a second resistance can readily be computed from the prescribed impedance.[38] The general theory of Part I can then be used to determine the corresponding sets of physically realizable open- and short-circuit impedances. It is easily shown that one of these

[35] This will be clarified by the subsequent description of the more detailed design procedure.

[36] In terms of the determinant Δ of a reactance network and various of its cofactors, the resistance part of the input impedance is represented by $\frac{R_2\Delta_{12}^2}{\Delta_{11}^2 - R_2^2\Delta_{1122}^2}$. It is obvious that single roots of the denominator can coincide with any of the double roots of the numerator only when R_2 has special values.

[37] Because the impedance function must generally be of less than normal degree for the configuration, there are usually equivalent 2-poles representing substantially more efficient use of the elements. In spite of this practical disadvantage, a knowledge of the canonical nature of this particular 2-pole is useful in general network theory problems.

[38] Specifically, the loss can be computed with the help of (16) of Part I.

will correspond to the prescribed impedance provided it is a positive real function.

Design Procedure

Consider now the actual operations involved in the design of a reactance 4-pole of the tandem section type producing prescribed open- and short-circuit impedances. As indicated above, the first step is to compute the input impedance of either end of the network corresponding to an arbitrary far-end termination subject only to the requirement that the degree of the impedance function must be normal for the open- and short-circuit impedances. A section of type A, B, C, or D is then designed in such a way that it will produce the required input impedance when terminated in a new physical impedance of lower degree. A second section is next designed to produce the required terminating impedance when itself terminated in a new impedance of further reduced degree. This procedure is continued until the required terminating impedance is reduced to a constant resistance.[39] Finally, the required terminating resistance is replaced by the equivalent combination of an ideal transformer terminated in a resistance identical with the termination assumed in computing the original input impedance.

Brune shows how any positive real function with a pole or root on the axis of real frequencies can be realized as the input impedance of a section of type A or B, respectively, terminated in an impedance of reduced degree. Brune also shows how any positive real function with no roots or poles on the real frequency axis but such that the resistance part has a pair of identical real or imaginary roots can be realized as the input impedance of a section of type C terminated in a physical impedance of reduced degree. Finally, if the roots of the resistance part of the input impedance all occur in identical pairs, the terminating impedance required in each case turns out to have this same property.[40] Hence, to complete the explanation of the design procedure it remains only to show how the appearance of identical complex roots of the resistance part permits a section of type D to be made use of in the same

[39] The exact manner in which the degree of the terminating impedance is reduced at each stage in the design is what leads to a final configuration for which the original input impedance is of normal degree.

[40] The computation of the terminating impedance from the input impedance may eliminate a pair of identical roots of the resistance part in reducing the degree of the impedance function and may convert one or more pairs of identical real roots into roots or poles of the impedance function or vice versa. Otherwise the roots of the resistance part remain unchanged.

way that a section of type C can be used when the roots are real or imaginary.

Design of Sections of Type D

Suppose that a prescribed section of type D produces a prescribed input impedance Z when terminated in impedance Z_L. If Δ is used to represent the determinant of the section of type D and if the input and output meshes are numbered 1 and 2, respectively, Z_L is related to Z by

$$Z_L = \frac{\Delta - Z\Delta_{11}}{Z\Delta_{1122} - \Delta_{22}} = \frac{\Delta_{12}^2 Z}{\Delta_{22}(\Delta_{22} - Z\Delta_{1122})} - Z_{s2} \qquad (22)$$

in which Z_{s2} represents the short-circuit impedance of the section of type D at the terminated end. If it is assumed that Z is a positive real function such that its resistance part has a pair of identical complex roots, it can be shown that the following condition always determines a physical section of type D which leads to a physical Z_L of lower degree than Z: The functions Δ_{12}^2 and $(\Delta_{22} - Z\Delta_{1122})$ are required to have coincident pairs of identical complex roots which are also coincident with a pair of identical complex roots of the resistance part of Z.

It is easily shown that the above coincidence of roots of Δ_{12}^2 and $(\Delta_{22} - Z\Delta_{1122})$ leads to the elimination of all roots of the impedance $(Z_L + Z_{s2})$ which are not also roots of Z, which excludes the possibility of roots in the right half of the p plane. If it is also assumed that the section of type D is physical, simple additional analysis shows that $(Z_L + Z_{s2})$ must actually be a positive real function. This proves the positive realness of Z_L itself except that it does not exclude real-frequency poles with negative residues covered up in $(Z_L + Z_{s2})$ by the positive residues of coincident poles of Z_{s2}. Finally, it can be shown that any coincident poles of the two impedances can be separated, without changing the corresponding residues of Z_L, by adding Lp to the original input impedance Z. The reduced degree of Z_L can be demonstrated by showing that the above conditions lead to eight coincidences of roots and poles of the expression $\dfrac{\Delta - Z\Delta_{11}}{Z\Delta_{1122} - \Delta_{22}}$ of (22).

To show that the condition on the roots of Δ_{12}^2 and $(\Delta_{22} - Z\Delta_{1122})$ actually leads to a physical section of type D, definite formulae for the element values are first derived from this condition without considering the question of physical realizability. The input impedance Z appears in these formulae only through the appearance of the values assumed

by Z and its derivative at double roots of its resistance part. This permits Z to be replaced in the formulae by the open-circuit impedance of a physical reactance 4-pole which produces the input impedance Z when terminated in the proper constant resistance.[41] It can then be shown that the corresponding capacities and self inductances of the section of type D will be finite real and positive.

Conditions Necessary for Ladder Networks

When all the frequencies of infinite loss are real or imaginary, sections of type D do not have to be included in the canonical tandem section configuration.[42] This is because the frequencies of infinite loss of the complete network are the corresponding critical frequencies of the separate sections, while sections of type D are required only to realize complex frequencies of infinite loss. When sections of type D are absent the network can be considered as a general ladder of reactances, which may be defined as any combination of alternate series and shunt branches consisting of reactance 2-poles. While the coupling between the series inductances in sections of type C renders them somewhat more complicated than alternate series and shunt 2-poles the coupling can be thought of as merely a device for realizing negative inductances appearing in the equivalent T networks.[43]

The Mid-Series Low-Pass Ladder Configuration

Of the large variety of actual configurations possible in ladders of the type described above, only a few are commonly made use of. An extensive special design theory has been developed for these particular configurations in order to permit the element values to be determined

[41] As was indicated previously, the theory of Part I can be used to demonstrate that any positive real function can be realized as the input impedance of a resistance-terminated reactance 4-pole.

[42] By the frequencies of infinite loss of a network is meant the frequencies of infinite loss obtained with general finite resistance terminations excluding any specific terminations which bring roots and poles of the general expression for the insertion voltage ratio into coincidence. Each frequency of infinite loss is included in one or more of the following groups of critical frequencies: roots of the open-circuit transfer impedance, poles of the short-circuit transfer impedance, coincident roots or poles of open- and short-circuit driving-point impedances of the same end of the network, and zeros of the resistance part of an input impedance obtained with a resistance termination.

[43] In many cases the negative inductance can be eliminated without the introduction of the coupling but the conditions under which this can be done are not subject to simple statement.

without the labor involved in using the general theory of tandem sections described previously. The special design theory is best developed first for the specific type of ladder indicated in Fig. 9, which may be referred to as the mid-series low-pass configuration even though it is not necessarily a mid-series type low-pass filter of the image parameter theory.[44] The other configurations commonly encountered can be designed by means of simple modifications of the theory of this special case.

The earliest special formulae for computing the element values of ladders of the mid-series low-pass configuration indicated in Fig. 9 were developed by Norton (7) as a part of his theory of constant resistance pairs of filters. Although Norton's formulae represented an important step in the development of the theory of mid-series low-pass ladders, it has been found that the computations which they call for in numerical problems are undesirably complicated and must usually be carried to an abnormally high precision.

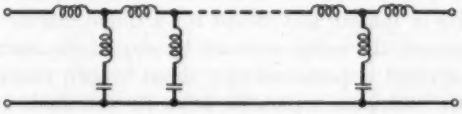


Fig. 9. The mid-series low-pass ladder configuration

As a result an extensive further analysis of the design problem has been carried out leading to the derivation of a new set of formulae. These new formulae are relatively satisfactory for numerical computations and are also useful in a variety of theoretical investigations, such as the determination of what impedances can be realized with mid-series low-pass ladders. In addition, the derivation of the formulae involves the development of an alternative set which are expressed very compactly in terms of determinants and which are useful in certain theoretical investigations even though they have the same disadvantages as Norton's when applied to ordinary numerical problems.

Assumptions and Conditions Leading to Design Formulae

The development of the design formulae for ladders of the mid-series low-pass configuration is simplified if certain simple assumptions are adopted temporarily. The procedure to be followed when the assumptions are not satisfied can best be investigated after the formulae have

[44] This type of ladder is obviously equivalent to sections of type C generally combined with one or more sections of type A consisting of simple series inductances.

been derived. In the first place, it is simplest to start by assuming a set of open- and short-circuit impedances to have been specified which are known in advance to be appropriate for the configuration, leaving until later the question of what impedances have this property. It is also best to assume further that the impedances are such that the multiplicity of solutions for the elements which are normally encountered are all possible, even though special sets of impedances can be found which require certain of the solutions to be excluded. Certain difficulties are also avoided by the temporary assumption that all the frequencies of infinite loss are different and that all the open- and short-circuit impedances are of normal degree for the configuration even though some may be of reduced degree in special cases.

The following relations, which can be shown to apply to any ladder of the mid-series low-pass configuration consistent with the above assumptions, form the basis of all known formulae for the element values.[45] First, the resonances of the shunt branches are identical with the frequencies of infinite loss except for a single infinite loss point at infinity.[46] Second, the value assumed by any of the open- and short-circuit driving-point impedances at a shunt branch resonance is independent of the elements separated from its terminals by the shunt branch, which acts as a short-circuit across the ladder. Finally, under the assumption of impedances of normal degree for the configuration the derivatives of the driving-point impedance functions with respect to the frequency have this same property.

It turns out that the relations stated above are sufficient to determine all element values from one open-circuit or short-circuit driving-point impedance together with the frequencies of infinite loss, except for the far end inductance in the case of an open-circuit impedance.[47] The

[45] These relations were introduced by Norton as the basis of his design equations.

[46] Under the present assumptions, the finite frequencies of infinite loss are the roots of the open-circuit transfer impedance and also the finite poles of the corresponding short-circuit impedance.

[47] This at first seems contrary to the well known fact that it takes three impedances to fix a 4-pole. The additional data are here supplied by the assumption of impedances appropriate for a specific configuration subject to special restrictions. Recall that Cauer's canonical reactance 4-pole of the shunt or admittance type can be designed from one short-circuit driving-point impedance and the short-circuit transfer impedance except for a two terminal shunt branch across the far end of the network. For the particular circuit under consideration the terminal shunt branch would be absent in the equivalent shunt type canonical network while the short-circuit transfer impedance could be found from a short-circuit driving-point impedance and the frequencies of infinite loss.

multiplicity of solutions which are normally obtainable are due only to the fact that the finite frequencies of infinite loss can be distributed arbitrarily among the shunt branches as their individual resonances. In order to obtain a unique solution it is expedient to assume at the outset that a particular distribution has been chosen. The problem then becomes that of realizing a known two-terminal impedance of the reactance type as an open-circuit or short-circuit impedance of a ladder of the midseries low-pass configuration with prescribed shunt branch resonances.

Continued Fraction Expansion Forming Basis of Design Problem

The development of the formulae for the element values calls for the introduction of extensive special notation so chosen as to reduce the design problem to the determination of a particularly simple continued fraction expansion of a known function. In the first place, instead of dealing directly with the element values, it is simpler to consider the

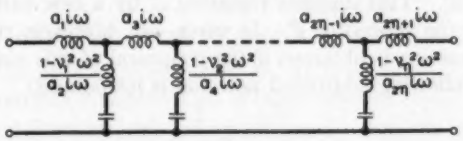


Fig. 10. Designation of the impedance branches of a mid-series low-pass ladder

constants in the designations indicated in Fig. 10 for the impedances of the various network branches, expressed in terms of ω rather than $p = i\omega$.[48] The constants $\nu_1, \cdots, \nu_\eta$ represent the reciprocals of the shunt branch resonances or finite points of infinite loss in terms of ω and are thus assumed to be known.[49] The problem under consideration therefore amounts to the determination of the so-called ladder network coefficients $a_1, \cdots, a_{2\eta+1}$, since there can be no difficulty in determining the element values from these coefficients together with the ν_σ's. In order to avoid ambiguity in the design formulae referred to previously, which are merely specific formulae for the individual a_σ's, it is convenient to include the additional requirement that the numbering of the coefficients a_1, a_2, etc., shall begin at the terminals of the particular

[48] Although the variable p is more convenient in formulating the general theory of Part I, ω turns out to be more convenient in the ladder network theory under consideration here.

[49] The reciprocals of the values of ω at infinite loss points are more convenient than the values themselves when the infinite loss points are later permitted to become infinite.

open-circuit or short-circuit driving-point impedance from which they are to be computed.

It is well known that the problem of realizing a 2-terminal impedance as the input impedance of a ladder network of prescribed configuration amounts to the problem of obtaining a prescribed form of continued fraction expansion of the impedance or of some related function.[50] For the particular network under consideration, the required continued fraction is simplest if the function F to be expanded is derived by dividing the impedance function by p, or by its equivalent $i\omega$. In other words, F is best defined by

$$F = \frac{Z}{i\omega} \tag{23}$$

where Z is the open- or short-circuit impedance from which the coefficients a_1, a_2, etc. are to be computed.

Since Z is an odd rational function of $i\omega$, the quantity F must be a function of ω^2. This suggests replacing ω^2 by a new variable in order to decrease the degree of F. It turns out, however, that a simpler continued fraction is obtained if the reciprocal of this variable is used. Hence the following additional notation is introduced:

$$z = \frac{1}{\omega^2} \tag{24}$$

$$z_\sigma = \nu_\sigma^2 \qquad \sigma = 1, \cdots, \eta$$

in which z represents the new variable, while the z_σ's indicate the values of z corresponding to the frequencies of infinite loss. In terms of this notation, the required continued fraction expansion of F takes the following form:

$$F = a_1 + \cfrac{1}{\cfrac{a_2}{z_1 - z} + \cfrac{1}{a_3 + \cfrac{1}{\cfrac{a_4}{z_2 - z} + \cdots}}} \tag{25}$$

The problem is to solve this identity for the a_σ's assuming the constants z_σ to be known and also the function F of the variable z.

In deriving the more useful formulae for the a_σ's which have been developed by solving the above problem the first part of the analysis

[50] A variety of ladders described by Fry (12), for instance, correspond to Stieltjes' Fractions.

is devoted to the derivation of the alternative formulae referred to previously as being expressed compactly in terms of determinants. The final formulae are then derived by expanding the determinants in terms of the partial fraction representation of the function F. Since the two parts of the derivation both involve long and complicated algebraic manipulation, they will be no more than briefly outlined here. Greater clarity will be obtained if the statement of the preliminary formulae in terms of determinants precedes the outline of their derivation.

Formulae for the a_σ's in Terms of Determinants

The determinants appearing in the preliminary solution for the a_σ's are formed from the quantities H_{qr} defined as follows in terms of the notation introduced above:

$$H_{qr} = \frac{F_q - F_r}{z_q - z_r} \qquad H_{qq} = F'_q \tag{26}$$

where F_q, F'_q, etc., are used to represent the values assumed by F and $\frac{dF}{dz}$ at $z = z_q$, etc. The use of the notation H_{qr} and H_{qq} is consistent since H_{qq} is the limit approached by H_{qr} as z_r approaches z_q.

The determinants themselves are of three different types. The determinant U_k is defined as $|\, H_{qr} \,|$ in which q and r take the values 1 to k. In other words,

$$U_k = \begin{vmatrix} H_{11} & H_{12} & \cdots & H_{1(k-1)} & H_{1k} \\ H_{21} & H_{22} & \cdots & H_{2(k-1)} & H_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{k1} & H_{k2} & \cdots & H_{k(k-1)} & H_{kk} \end{vmatrix} \tag{27}$$

The determinant V_k is obtained from U_k by changing the elements of the last column to unity. In other words,

$$V_k = \begin{vmatrix} H_{11} & H_{12} & \cdots & H_{1(k-1)} & 1 \\ H_{21} & H_{22} & \cdots & H_{2(k-1)} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{k1} & H_{k2} & \cdots & H_{k(k-1)} & 1 \end{vmatrix} \tag{28}$$

Finally, the determinant W_k is obtained from U_k by changing the elements of the last column from $H_{\sigma k}$ to $H_{\sigma(k+1)}$. Thus,

$$W_k = \begin{vmatrix} H_{11} & H_{12} & \cdots & H_{1(k-1)} & H_{1(k+1)} \\ H_{21} & H_{22} & \cdots & H_{2(k-1)} & H_{2(k+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{k1} & H_{k2} & \cdots & H_{k(k-1)} & H_{k(k+1)} \end{vmatrix} \tag{29}$$

In terms of these determinants all the ladder network coefficients except for $a_{2\eta+1}$ are given by the following design equations:[51]

$$\begin{aligned} a_1 &= F_1 & a_{2k-1} &= \frac{(z_k - z_{k-1})U_{(k-1)}W_{(k-1)}}{V_{(k-1)}V_k} \\ a_2 &= \frac{-1}{U_1} & a_{2k} &= \frac{-V_k^2}{U_{(k-1)}U_k} \end{aligned} \tag{30}$$

$$k = 2 \cdots \eta$$

The coefficient $a_{2\eta+1}$ corresponds to the inductance forming the last series branch. If the function F is obtained from an open-circuit impedance, the coefficient must be determined from some other impedance. If F corresponds to a short-circuit impedance, $a_{2\eta+1}$ can be found from the value assumed by F at zero frequency, or by using the above formulae with an arbitrary additional constant $z_{\eta+1}$.[52]

The above formulae can be derived by a method of induction which is reasonably straightforward though long and tedious. The formulae for the first four coefficients are first derived from Norton's equations or directly from the behavior of the continued fraction (25) in the neighborhood of $z = z_1$ and $z = z_2$. This yields the special formulae for a_1 and a_2 indicated above and shows the formulae for a_3 and a_4 to be consistent with the general formulae indicated for a_{2k-1} and a_{2k}. The derivation is then completed by showing that if the formulae for a_{2k-1} and a_{2k} are correct, then those for a_{2k+1} and a_{2k+2} must also be correct. This is demonstrated by first using the formulae for a_{2k-1} and a_{2k} to express a_{2k+1} and a_{2k+2} in terms of the impedance that would be obtained by removing the first two branches of the network, corresponding to

[51] The formulae for a_2, a_3, a_4, involve U_1, V_1, W_1. As defined by (27), (28) and (29) these quantities appear strange in that they are determinants of only the first order. These first order determinants, however, merely represent H_{11}, 1, H_{12}, respectively.

[52] If F corresponds to a short-circuit impedance, an arbitrary shunt branch can be assumed to be connected across the short-circuit since it will not affect F. This permits $a_{2\eta+1}$ to be determined exactly as though there were a complete additional "section."

a_1 and a_2. This "reduced" impedance is then replaced by an equivalent expression in terms of the original impedance and the formulae for a_1 and a_2. Considerable manipulation of the determinants in the resulting equations finally yields the general formulae for a_{2k+1} and a_{2k+2}.

The formulae (30) are indeterminate unless all the constants $z_1, \cdots, z_q$ are different in accordance with the original assumption of no two identical frequencies of infinite loss. Coincident frequencies of infinite loss can be handled, however, by assuming infinitesimal differences and making use of a Taylor's series expansion of the function F representing $\frac{Z}{i\omega}$ considered as a function of $z = \frac{1}{\omega^2}$. When all the frequencies of infinite loss are identical except for the single one at infinity, the continued fraction (25) becomes a Stieltjes' fraction of a type considered by Fry (12). The known formulae for the constants in the Stieltjes' fraction expansion are undoubtedly derivable from (30) by the Taylor's series method.[53]

The effect of abandoning the original assumptions other than that of no two identical frequencies of infinite loss can best be considered after the final formulae have been developed by showing how the determinants of the first set can be expanded.

Derivation of the Final Design Formulae by Expanding the Determinants

It is readily shown that if the function F is expanded into a sum of partial fractions, this expansion will always take the following form provided the open- or short-circuit driving-point impedance from which it is derived is physically realizable:

$$F = -B_0 z + B_\infty + \frac{B_1}{z - \beta_1} + \cdots + \frac{B_\mu}{z - \beta_\mu} \tag{31}$$

In this expression the B's and β's are all positive and are also all finite except that B_0 may sometimes be zero.

In terms of the partial fraction expansion of F, the determinant element H_{qr} as defined in (26) becomes

$$H_{qr} = -B_0 - \frac{B_1}{(\beta_1 - z_q)(\beta_1 - z_r)} - \cdots - \frac{B_\mu}{(\beta_\mu - z_q)(\beta_\mu - z_r)} \tag{32}$$

[53] This derivation has not been proved rigorously but has been carried far enough to indicate the way in which the transformation of the formulae takes place.

while H_{qq} is obtained by merely equating z_r to z_q in this formula. The determinants U_k, V_k, W_k defined in (27) through (29) and appearing in the formulae (30) for the a_s's can be expanded in terms of these partial fraction representations of the H_{qr}'s. The derivation of these expansions, however, is too long and complicated to be more than briefly outlined here.

In deriving the expansions of the determinants, the particular W_k determinant of degree identical with the number of partial fractions in the expansion of H_{qr} is first examined. It is found that this particular W_k determinant can be expressed in terms of a product of two determinants of the general form $\left|\frac{1}{x_i - y_j}\right|$, which are evaluated in well-known treatises on determinant theory.[54] It is then shown that W_k determinants of higher degree must vanish while those of lower degree are equivalent to sums of similar factorable determinants.[55] Each term in these sums is actually the determinant that would be obtained by using only k of the partial fractions of H_{qr}, i.e., by setting all but k of the B's in (32) equal to zero. There must be one such term for every possible choice of k partial fractions. After the expansions of the W_k determinants have been determined the U_k and V_k determinants can be expanded by treating them as certain limiting cases of W_k determinants.

When the expansions of the determinants are inserted in the equations (30) for the a_s's, a variety of factors in differences between the infinite loss points z_1, z_2, etc., can be cancelled out. The formulae then take the form

$$\begin{aligned} a_1 &= F_1 & a_{2k-1} &= \frac{\Psi_{U(k-1)}\Psi_{W(k-1)}}{\Psi_{V(k-1)}\Psi_{Vk}} \\ a_2 &= \frac{1}{\Psi_{U1}} & a_{2k} &= \frac{\Psi_{Vk}^2}{\Psi_{U(k-1)}\Psi_{Uk}} \end{aligned} \tag{33}$$

in which the Ψ's represent the uncancelled parts of the expansions of the original determinants.

The Ψ's themselves are best expressed in terms of the quantities u_{qk}, v_{qk}, w_{qk} defined by the recursion formulae

[54] See, for instance, the chapter on functional determinants in the treatise by Scott and Mathews (13).

[55] The similar vanishing of higher degree determinants of the U_k and V_k types shows the finite nature of the continued fraction (25).

$$v_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)}$$

$$u_{q0} = B_q \qquad u_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)^2} \tag{34}$$

$$w_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)(\beta_q - z_{k+1})}$$

The statement of the Ψ's in terms of these quantities requires a very complicated summation and product notation in the general case. Hence, it will be better to avoid the necessity of a statement of the general case by listing enough specific cases to establish what such a statement would have to show.

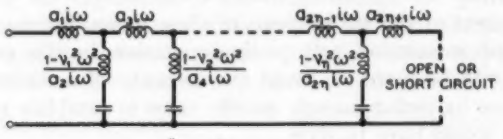
The formulae for the simpler Ψ's are listed in Table I together with the additional previous relations necessary in the actual computation of ladder network coefficients. These are sufficient to indicate the general case and are also sufficient by themselves for ordinary design purposes, particularly when impedances corresponding to both ends of a network are known so that part of the elements can be determined from each end. Some additional simplifications can be obtained, however, by developing more specialized forms of the equations for specific numbers of sections. The behavior of the impedances at zero and infinity can also be used advantageously in determining one or two coefficients for which the standard formulae are most complicated or for checking purposes. Special simplified formulae have also been derived for the ladder equivalents of symmetrical lattice networks, which are normally computed first in the design of symmetrical circuits. These formulae can best be introduced at a later point, however, when the theory of lattice networks is considered.

In ordinary numerical problems, the expanded formulae do not require the extremely high precision of computation which is commonly necessary when Norton's equations are used or the formulae in terms of determinants. Similarly, the expanded formulae do not become indeterminate when coincident frequencies of infinite loss are encountered, as do the other formulae. They also lead to somewhat more straightforward numerical computations in ordinary design problems, although the extent to which this is true may vary widely from problem to problem. While the complexity of the formulae increases so rapidly with network complexity as to render them unsatisfactory for the design of ladders of more than four or possibly five sections, such complicated networks are rarely encountered. Finally, the formulae for ladders of

general numbers of sections assume a form which renders them useful in such general studies as the investigation of the requirements on impedances of low-pass ladders or of the possibility of avoiding coupled coils, etc.

TABLE I

General Mid-Series Low-Pass Ladders



$$F = \frac{Z}{i\omega} = -B_0 z + B_\infty + \sum_{q=1}^{\mu} \frac{B_q}{z - \beta_q} \qquad z = \frac{1}{\omega^2} \qquad \alpha = \infty \text{ at } z = z_k = \nu_k^2$$

$$u_{q0} = B_q \qquad v_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)} \qquad u_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)^2} \qquad w_{qk} = \frac{u_{q(k-1)}}{(\beta_q - z_k)(\beta_q - z_{k+1})}$$

$$\Psi_{V1} = 1 \qquad \Psi_{V2} = \sum v_{q2} \qquad \Psi_{V3} = \sum v_{q3} v_{r3} (\beta_q - \beta_r)^2$$
$$\Psi_{V4} = \sum v_{q4} v_{r4} v_{s4} (\beta_q - \beta_r)^2 (\beta_q - \beta_s)^2 (\beta_r - \beta_s)^2$$

$$\Psi_{U1} = \sum u_{q1} + B_0 \qquad \Psi_{U2} = \sum u_{q2} u_{r2} (\beta_q - \beta_r)^2 + B_0 \sum u_{q2}$$
$$\Psi_{U3} = \sum u_{q3} u_{r3} u_{s3} (\beta_q - \beta_r)^2 (\beta_q - \beta_s)^2 (\beta_r - \beta_s)^2 + B_0 \sum u_{q3} u_{r3} (\beta_q - \beta_r)^2$$

$$\Psi_{W1} = \sum w_{q1} + B_0 \qquad \Psi_{W2} = \sum w_{q2} w_{r2} (\beta_q - \beta_r)^2 + B_0 \sum w_{q2}$$
$$\Psi_{W3} = \sum w_{q3} w_{r3} w_{s3} (\beta_q - \beta_r)^2 (\beta_q - \beta_s)^2 (\beta_r - \beta_s)^2 + B_0 \sum w_{q3} w_{r3} (\beta_q - \beta_r)^2$$

$$\sum u_q u_r = \sum_{q=1}^{\mu-1} \sum_{r=q+1}^{r=\mu} u_q u_r = u_1 u_2 + u_1 u_3 + u_2 u_3 + \cdots + u_\mu u_{\mu-1}$$

and similarly for the other sums of products of terms.

$$a_1 = B_\infty - B_0 z_1 - \sum v_{q1} \qquad a_{2k-1} = \frac{\Psi_{U(k-1)} \Psi_{W(k-1)}}{\Psi_{V(k-1)} \Psi_{Vk}}$$

$$a_2 = \frac{1}{\Psi_{U1}} \qquad a_{2k} = \frac{\Psi_{Vk}^2}{\Psi_{U(k-1)} \Psi_{Uk}}$$

$$k = 2, \cdots, \eta$$

Insertion Loss Functions and Impedances Realizable with Ladders of the Mid-Series Low-Pass Configuration

The mid-series low-pass ladder configuration turns out to be appropriate for the realization of insertion power ratios of the form

$$e^{2\alpha} = \frac{1 + \Gamma_1 \omega^2 + \Gamma_2 \omega^4 + \cdots + \Gamma_{2\eta+1} \omega^{4\eta+2}}{(1 - \nu_1^2 \omega^2)^2 (1 - \nu_2^2 \omega^2)^2 \cdots (1 - \nu_\eta^2 \omega^2)^2} \tag{35}$$

in which η represents the number of shunt branches in the ladder while the Γ_σ's and ν_σ's are arbitrary constants. Corresponding open- and short-circuit impedances realizable with mid-series low-pass ladders and subject to the previous assumptions of normal degrees for the configuration and of normal multiplicity of solutions for the element values can always be found by a straightforward method except in certain special cases corresponding to discrete choices of the Γ_σ's and ν_σ's.[56]

The first step in determining the impedances is to find a solution for the polynomials A, B, A', and B' of the general impedance theory of Part I, using the numerator and denominator of (35) as the polynomials N and P^2. It is readily shown that the multiplicity of solutions for the polynomials, which was indicated in part I, is such that the signs of A' and B' can be chosen arbitrarily as far as general realizability is concerned and also the sign of $(A + pB)$. Except in the very special cases referred to above, the impedance formulae exhibited in equations (18) in Part I yield corresponding impedances realizable with ladders of the desired type provided the signs of A', B', and $(A + pB)$ are chosen in accordance with the following conditions: $\frac{A}{P}$ must be positive at zero frequency, $\frac{B'}{B}$ must be negative at infinite frequency, and $\frac{A'}{A}$ must be positive or negative at zero frequency depending upon whether termination R_1 is greater than or less than R_2.[57]

Difficulties Encountered in Special Cases

The difficulties which can be encountered in special cases are of two types. The first type can be encountered even though the impedance functions are of the form normal for the configuration. The second

[56] By discrete choices of the constants is meant choices which can always be avoided by small changes in a single constant. It is assumed, of course, that $e^{2\alpha}$ meets the physical requirement that it must be positive and no less than $\frac{4R_1R_2}{(R_1+R_2)^2}$ at all real frequencies.

[57] Actually all solutions are normally realizable if modified mid-series low-pass ladders are permitted. The requirement upon the sign of $\frac{A}{P}$ or $\frac{A'}{A}$ can be violated, for instance, if an ideal transformer is included in one end of the network. Similarly, the requirement upon the sign of $\frac{B'}{B}$ can normally be violated if a shunt condenser is added to one end of the ladder and if negative series inductances realizable with perfectly coupled coils are permitted.

type corresponds to the appearance of various open- or short-circuit impedances which are of reduced degree because of the coincidence of roots of the numerators and denominators of the corresponding general formulae.

It can be shown that if all the impedances are of normal degree, no finite frequency of infinite loss can coincide with a root or pole of an open- or short-circuit driving-point impedance. It follows that the corresponding values of the quantities v_{qk}, u_{qk}, w_{qk} appearing in the design formulae of Table I will all be finite. Of these quantities, u_{qk} will always be positive but v_{qk} and w_{qk} can be negative. As a result, the quantities Ψ_{Uk} formed from the u_{qk}'s in the manner indicated in Table I will all be finite but the quantities Ψ_{Vk} and Ψ_{Wk} may be zero. The vanishing of Ψ_{Wk} merely replaces a series inductance by a simple conductor but the vanishing of Ψ_{Vk} leads to the requirement of three network branches with infinite impedances at all frequencies, one shunt branch and the two adjacent series branches.[55]

If it is no longer assumed that all the multiplicity of solutions for the a_σ's are to be physical, difficulties of this type can normally be overcome by modifying the choice of the particular frequencies of infinite loss which are to be the resonances of the individual shunt branches or by choosing a different set of impedances corresponding to the same insertion loss. It is within the bounds of possibility, however, to encounter cases in which all solutions lead to the same difficulties. The ladder must then be modified to the extent of using at least one anti-resonant circuit as a series branch. This can be done by modifying properly the normal design procedure.

When there are finite frequencies of infinite loss which are coincident with roots or poles of open- or short-circuit driving-point impedances, some of the impedance functions will be of reduced degree. It is then normally possible to realize the impedances by adding terminal series or shunt branches or both to a mid-series loss-pass ladder with impedances of normal degree for its configuration. It is also usually possible, however, to obtain complete networks of the normal mid-series low-pass form by merely using impedances that are still of normal degree in computing the element values. When all impedances are of reduced degree, modifications of the normal design procedure can be used or else the general method applying to the canonical tandem section configuration.

[55] Ψ_{vk}, Ψ_{wk} may also be negative, of course, leading to negative elements, but these turn out to be realizable with coupled coils.

Elimination of Various Elements

There is no difficulty in permitting any or all of the frequencies of infinite loss to be placed at infinity. This merely requires the corresponding shunt branches to be simple condensers rather than resonant circuits. The design is carried out by setting the proper ν_σ's equal to zero in (35) and also the corresponding z_σ's representing their squares in the design formulae of Table I.

Certain power ratios of the general form of (35) also lead to the vanishing of one or more of the series inductances. One special case of this type is of particular importance. In this case, one terminal series branch vanishes while the next shunt branch is a simple capacity, which leaves a network of the type of Fig. 11. The proper form of power ratio is obtained from (35) by reducing the terms in the numerator by one and also the number of ν_σ's.[59] In other words,

$$e^{2\alpha} = \frac{1 + \Gamma_1\omega^2 + \Gamma_2\omega^4 + \cdots + \Gamma_{2\eta}\omega^{4\eta}}{(1 - \nu_1^2\omega^2)^2(1 - \nu_2^2\omega^2)^2 \cdots (1 - \nu_{\eta-1}^2\omega^2)^2} \tag{36}$$

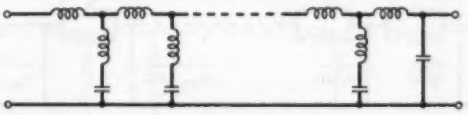


Fig. 11. A special form of mid-series low-pass ladder configuration corresponding to a special form of insertion loss function

Other Types of Ladders Commonly Encountered

Ladders of the configuration of Fig. 12, which may be described as the mid-shunt low-pass configuration, are most easily designed by redefining the coefficients $a_1, \cdots, a_{2\eta+1}$ in accordance with the designations of the branch impedances indicated in the figure. These coefficients are related to the short- and open-circuit admittances in exactly the same way that the coefficients of ladders of the mid-series type are related to the open- and short-circuit impedances. As a result, mid-series and mid-shunt ladders will produce the same insertion losses when terminated in R_1 and R_2 provided their coefficients are related by the equations

$$a_{2\sigma-1} = R_1R_2\bar{a}_{2\sigma-1} \qquad \bar{a}_{2\sigma} = R_1R_2a_{2\sigma} \tag{37}$$

[59] The vanishing of a terminal series inductance next to a resonant shunt branch, however, does not normally change the degree of the numerator in the power ratio expression in this way.

in which the a_σ's and $\bar{a}_\sigma$'s are the coefficients of the mid-series and mid-shunt ladders, respectively, and must be numbered from opposite ends in the two networks.

Negative inductances realizable with coupling in series type ladders become negative capacities in the corresponding shunt type configurations. These can be realized, however, by introducing ideal transformers in the proper way. To understand how this can be accomplished it is only necessary to note that an ideal transformer shunted by any 2-pole, such as a condenser, is equivalent to a pair of "perfectly coupled impedances" exactly similar to perfectly coupled inductances.

Other types of ladders can be designed by making use of the well-known method of frequency transformations. The previous design formulae apply to the determination of a low-pass ladder of reactance elements producing an insertion loss represented by an appropriate function of ω. Suppose some other insertion loss function is transformed

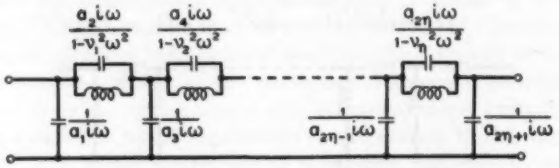


Fig. 12. The mid-shunt low-pass ladder configuration

into this same function by replacing ω by a related variable Ω. Since the reactance elements of the low-pass filter theory are merely devices for producing impedances proportional to $i\omega$ or its reciprocal, the same theory can now be used to design a corresponding ladder made up of impedances proportional to $i\Omega$ or its reciprocal. If all the elements of the original low-pass ladder are positive, the impedance branches of the transformed ladder can be realized with physical 2-poles provided $i\Omega$ represents a physical impedance function. If the original low-pass ladder includes negative inductances realizable with coupled coils, the negative elements in the transformed ladder can be realized by using ideal transformers just as in mid-shunt low-pass ladders including negative capacities.

The mid-series and mid-shunt high-pass configurations obtained by replacing inductances by capacities and vice versa in low-pass ladders can be designed by defining Ω as $-\frac{1}{\omega}$ and using power ratios represented

by functions of Ω identical with the functions of ω appropriate for the low-pass configurations. Band-pass configurations can be designed by defining Ω as $\frac{\omega^2 - \omega_m^2}{\omega}$ provided their insertion loss characteristics when plotted against log (ω) are to be symmetrical about log (ω_m). They can then be realized as combinations of series and parallel resonant circuits all resonating at ω_m. The situation is exactly similar in regard to band-elimination configurations, for which Ω is $\frac{-\omega}{\omega^2 - \omega_m^2}$.[60]

The only other ladder configurations commonly encountered are the more general band-pass type indicated in Fig. 13, and its inverse. The series type illustrated can normally be designed as an equivalent simpler

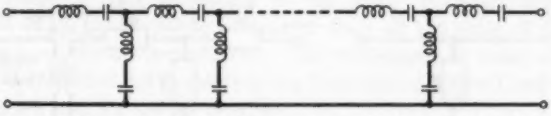


Fig. 13. The general mid-series band-pass ladder configuration

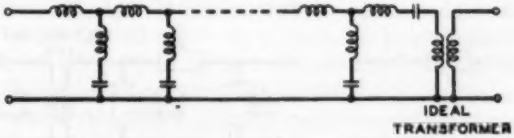


Fig. 14. An equivalent of the network of Fig. 13

network of the configuration indicated in Fig. 14.[61] It can be determined from the equivalent network by means of the so-called impedance

[60] It is sometimes more convenient, of course, to include arbitrary constant factors in the frequency transformations. In the design of a band-pass filter, for instance, it is frequently convenient to start by designing a low-pass filter with a cut-off at $\omega = 1$. Then the desired band-pass filter is obtained by replacing ω by an Ω representing

$$\frac{\omega^2 - \omega_{c1}\omega_{c2}}{\omega(\omega_{c2} - \omega_{c1})},$$

where ω_{c1}, ω_{c2} correspond to the two cut-offs of the band-pass filter.

[61] An equivalent of this type always exists except in limiting cases in which one or more shunt branches are simple inductances. A similar equivalent configuration obtained by replacing inductances by capacities and vice versa always exists except when one or more shunt branches are simple condensers. When simple inductance and capacity shunt branches are both encountered, special design methods must be used.

transformation indicated in Fig. 15A.[62] The equivalent network itself is of such a form that the formulae of Table I can be applied directly to its design. The short-circuit driving-point impedance measured at the terminals farther from the transformer, for instance, is determined by a configuration exactly the same as that determining an open-circuit impedance of a mid-series low-pass ladder with a far-end terminal shunt branch consisting of a simple condenser. The only operation not covered by the formulae of Table I is the determination of the impedance ratio of the transformer, which turns out to be determined by the behavior of the open- and short-circuit driving point impedances at zero frequency.

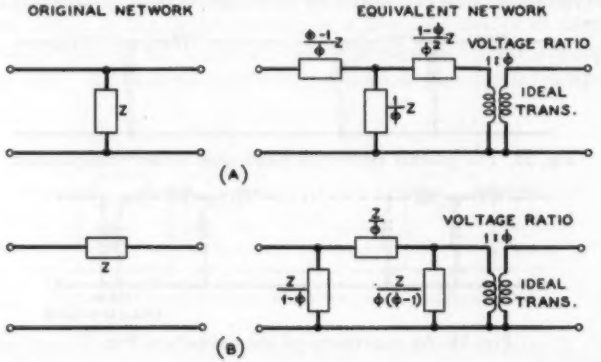


Fig. 15. Network equivalences indicating the impedance transformation principle

The configuration inverse to that of Fig. 13, which is made up of parallel resonant circuits in place of series resonant circuits, is designed by an exactly similar method with the help of the equivalence indicated in Fig. 15B.

Sufficient Conditions for Positive Elements

From the standpoint of actual construction, the perfectly coupled inductances and ideal transformers which are sometimes required for the realization of negative elements in physical ladders are highly undesirable. The two conditions stated below are sufficient to insure that

[62] The impedance transformation principle represented by Figs. 15A and 15B was discovered by Norton (14). In the determination of a network of the configuration indicated in Fig. 13 from one of the type indicated in Fig. 14 there is a considerable arbitrariness as to how the series capacity is distributed among the different series branches.

all the elements in a mid-series low-pass ladder will be positive, making unnecessary the use of coupled coils or ideal transformers in the ladder itself or in others related to it by frequency transformations or inverse relationships such as (37). While these conditions are not necessary, they turn out to be useful in demonstrating that most ladders encountered in filter design can be expected to be realizable without the use of coupling.

Assuming a mid-series low-pass configuration and impedances of the appropriate general form the first condition calls for frequencies of infinite loss which are all real (or infinite) and are greater than all the finite poles of at least one of the open-circuit driving-point impedances. The second condition requires the particular frequencies of infinite loss corresponding to the resonances of the shunt branches nearest the terminals to be also equal to or greater than all the roots of the corresponding open-circuit impedances. The sufficiency of these conditions is easily established by examining the formulae of Table I and recalling that the B's and β's are all positive.

One interesting special case is that in which all frequencies of infinite loss occur at infinity. It is seen that the above conditions are always satisfied in this special case. As a result, all power ratios of the form (35) with all ν_σ's reduced to zero are realizable with networks of the same *configuration* as constant-k low-pass image parameter filters provided $e^{2\alpha}$ meets the general physical requirement that it must be no less than $\frac{4R_1R_2}{(R_1+R_2)^2}$ at all real frequencies. The same situation also holds in regard to the power ratio (36) with all ν_σ's reduced to zero, the corresponding constant-k configuration merely including an odd number of "half sections." These two special power ratios include all even polynomials in ω which have unit constant terms and which satisfy the physical limit on $e^{2\alpha}$ at real frequencies.

Symmetrical and Inverse Impedance Ladders

The ladders used as filters usually have impedances and terminations meeting one of two special conditions.[63] One condition calls for an electrically symmetrical network and equal terminations. The other requires each open-circuit impedance to be the inverse of the short-circuit impedance of the other end of the network with respect to the mean of the terminations.[64]

[63] Except when they are modified to compensate for effects of dissipation.

[64] If the image impedances and transfer constants are used in the description of the networks, these conditions require equal and inverse image impedances, respectively.

If the requirement of symmetry or of inverse impedances is to be satisfied, the polynomial A' or B', respectively, must be identically zero in the formulae (18) of Part I relating the impedances to the insertion loss. An examination of the relation of A' and B' to the insertion power ratio shows that their vanishing requires expressions of the form

$$e^{2\alpha} = 1 + \left(\frac{\omega B'}{P}\right)^2 \qquad (38)$$

$$A' = 0$$

in the case of symmetrical networks and equal terminations, and of the form

$$e^{2\alpha} = \frac{4R_1R_2}{(R_1 + R_2)^2} + \left(\frac{A'}{P}\right)^2 \qquad (39)$$

$$B' = 0$$

in the case of inverse impedance networks.

The specification of power ratios in the forms (38) and (39) simplifies the design procedure in that no roots need be extracted in determining A' and B', the only root extraction being that involved in finding the polynomials A and B. It is readily shown that the conditions necessary for the physical realizability of these power ratios permit A', B', and P to be any even polynomials in ω with real coefficients. In other words, there will be at least one corresponding symmetrical or inverse impedance network for every power ratio of the general form

$$e^{2\alpha} = \Lambda(1 + \Phi^2) \qquad (40)$$

in which Φ is any odd or even rational function of ω with real coefficients while Λ represents $\frac{4R_1R_2}{(R_1 + R_2)^2}$ and must be unity when Φ is an odd function.[65]

[65] When Φ is odd and Λ is different from unity, the ratio of the open-circuit driving-point impedances, which is identical with the ratio of the corresponding short-circuit impedances, will be equal to the ratio of the terminations. This permits the use of a symmetrical network combined with an ideal transformer. It is convenient to permit unequal terminations in the case of inverse impedance networks and not in the case of symmetry, or of proportional impedances, in that unequal terminations are usually necessary in the inverse impedance case if ideal transformers are to be avoided.

When all the poles of $\frac{B'}{P}$ are real or imaginary and are also finite, the power ratio (38) leading to symmetrical networks can be expressed in the form (35) appropriate for mid-series low-pass ladders of the normal configuration indicated in Fig. 9. When $\frac{A'}{P}$ obeys the same condition except for a single pole in terms of ω^2 at infinity and when also R_1 and R_2 are so chosen that α is zero at zero frequency, the power ratio (39) leading to inverse impedance networks can be expressed in the form (36) appropriate for mid-series low-pass ladders terminated at one end in shunt condensers as in Fig. 11. Exactly similar relations exist in regard to ladders of other than the mid-series low-pass configuration.

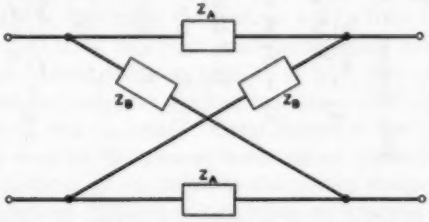


Fig. 16. The balanced lattice configuration

Lattice Networks and the Design of Symmetrical Ladders

It is well known that there is an equivalent lattice network of the type indicated in Fig. 16 corresponding to every physical network which is electrically symmetrical.[66] When the open- and short-circuit impedances are known, the impedance arms Z_A, Z_B can be very easily computed by well-known formulae. When a lattice of reactances is to be designed to produce a power ratio prescribed in the form (38), however, it is simpler to determine the impedance branches by the formulae described below rather than to first compute the open- and short-circuit impedances and then the impedance arms.

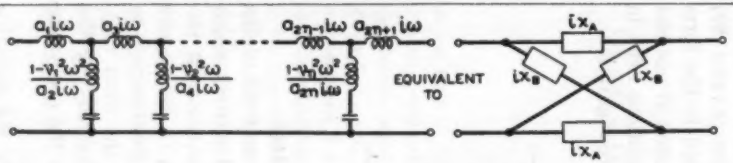
When a power ratio is prescribed in the form (38), the quantity $(P + pB')$ can obviously be expressed as the following product of two polynomials in p:

$$P + pB' = (P_1 + pB_1)(P_2 - pB_2) \tag{41}$$

[66] This was pointed out by Campbell (15).

TABLE II

Ladder Equivalents of Mid-Series Low-Pass Lattices



One Section — $\eta = 1$

$$X_A = \gamma_1\omega \qquad X_B = \frac{\gamma_2\omega^2 - \gamma_3}{\omega}$$

$X_A = X_B$ at infinite loss point $\frac{1}{\omega^2} = z_1$

$$a_1 = \gamma_1 \qquad a_2 = \gamma_1$$

$$a_3 = \frac{2}{\gamma_2}$$

Two Sections — $\eta = 2$

$$X_A = \frac{\omega(\gamma_1\omega^2 - \gamma_2)}{(\beta_1\omega^2 - 1)} \qquad X_B = \frac{\gamma_2\omega^2 - \gamma_4}{\omega}$$

$X_A = X_B$ at infinite loss points $\frac{1}{\omega^2} = z_1 , z_2$

$$a_1 = (\gamma_2 - \gamma_4 z_1) \qquad a_4 = \frac{2(\beta_1 - z_2)}{\gamma_4(2\beta_1 - z_1 - z_2)}$$

$$a_2 = \frac{2(\beta_1 - z_1)}{\gamma_4(2\beta_1 - z_1 - z_2)} \qquad a_5 = (\gamma_2 - \gamma_4 z_2)$$

$$a_3 = \gamma_4(2\beta_1 - z_1 - z_2)$$

Three Sections — $\eta = 3$

$$X_A = \frac{\omega(\gamma_1\omega^2 - \gamma_2)}{(\beta_1\omega^2 - 1)} \qquad X_B = \frac{(\gamma_2\omega^4 - \gamma_4\omega^2 + \gamma_6)}{\omega(\beta_2\omega^2 - 1)}$$

$X_A = X_B$ at infinite loss points $\frac{1}{\omega^2} = z_1 , z_2 , z_3$

$$T_1 = 2\beta_1\beta_2 - (\beta_1 + \beta_2)(z_1 + z_2) + 2z_1z_2$$

$$T_2 = 2(\beta_1 - z_3)(\beta_1 - z_1)(\beta_2 - z_1) + (z_1 - z_2)T_1$$

$$T_3 = 2(\beta_1 - z_3)(\beta_1 - z_1)(\beta_2 - z_3) + (z_3 - z_2)T_1$$

$$T_4 = \frac{(\gamma_2\beta_1 - \gamma_1)}{(\beta_1 - z_1)(\beta_1 - z_2)(\beta_1 - z_3)} = \frac{\gamma_6}{(\beta_1 - \beta_2)}$$

$$a_1 = \frac{\gamma_1 - \gamma_2 z_1}{\beta_1 - z_1} \qquad a_5 = \frac{T_2T_4(\beta_1 - z_1)}{T_1}$$

$$a_2 = \frac{2(\beta_1 - z_1)(\beta_2 - z_1)}{T_2T_4} \qquad a_6 = \frac{2(\beta_1 - z_3)(\beta_2 - z_3)}{T_3T_4}$$

$$a_3 = \frac{T_2T_4(\beta_1 - z_3)}{T_1} \qquad a_7 = \frac{\gamma_1 - \gamma_2 z_3}{\beta_1 - z_3}$$

$$a_4 = \frac{2(\beta_1 - z_2)(\beta_2 - z_2)T_1^2}{T_2T_3T_4(\beta_1 - \beta_2)}$$

where P_1, B_1, P_2, B_2 are even polynomials such that the roots of $(P_1 + pB_1)$ are all those roots of $(P + pB')$ which have negative real parts. It turns out that the impedance arms Z_A and Z_B of one lattice producing the prescribed power ratio are related to P_1, B_1, etc., by the formulae

$$Z_A = R\frac{pB_1}{P_1} \qquad Z_B = R\frac{P_2}{pB_2} \tag{42}$$

in which R represents the equal terminating resistances. The only three other lattices of reactances corresponding to the prescribed power ratio are obtained by interchanging these impedances and replacing them by their inverses with respect to R.[67]

In the design of symmetrical ladders with prescribed insertion losses, it is usually easier to determine the element values from the equivalent lattices rather than from the open- and short-circuit impedances. In the design of mid-series low-pass ladders of one, two or three shunt branches or of related networks, for instance, the special design formulae listed in Table II can be used. These formulae can be derived in much the same way as the general formulae of Table I by using an open-circuit impedance in formulating the partial fraction expansions of the determinants in (30) and noting relations between the constants required for symmetry. The relations between the constants are due to the fact that the open-circuit impedance is proportional to the sum of the two impedance arms, which requires the sums of the corresponding partial fractions to be equal at frequencies of infinite loss. Even when no other special relations are used, the initial determination of the open-circuit impedance as a sum of two functions simplifies the computation of the partial fraction expansion necessary for the use of Table I.

Part III. Special Insertion Loss Functions for Filtering Purposes

From the standpoint of filter design the general theory of Parts I and II is incomplete in two principal respects. In the first place, it gives no indication as to how to choose general insertion loss functions in such a way as to obtain efficient filter characteristics. In addition, the general design procedure is extremely complicated in numerical

[67] The polynomials B' and P are somewhat arbitrary as far as the power ratio of the form (38) is concerned. It turns out, however, that there are only four corresponding lattices in spite of this arbitrariness.

problems, involving the determination of the roots of two high degree polynomials. Part III rectifies this situation by introducing special types of realizable insertion loss functions which represent efficient filter characteristics and which also lead to relatively simple special design procedures.

Assumption of Insertion Loss Characteristics Appropriate for Symmetrical or Inverse Impedance Networks

The special insertion loss functions used for filtering purposes are appropriate for the symmetrical and inverse impedance networks described in the closing paragraphs of Part II. It turns out that any loss function representing an efficient filter characteristic at least approximates one of the types required for these particular networks. In addition, as was indicated previously, these networks can be designed with only half the root extraction which is more generally required.

Recall that insertion power ratios which are to be appropriate for symmetrical networks with equal terminations or for networks with impedances which are inverse with respect to the mean of the terminations must be of the form

$$e^{2a} = \Lambda(1 + \Phi^2) \tag{43}$$

where Φ is any odd or even rational function of frequency in the two cases, respectively, and Λ represents $\frac{4R_1R_2}{(R_1 + R_2)^2}$ (being thus unity in the case of symmetrical networks with equal terminations). It is obvious that the poles of Φ are frequencies of infinite loss while the roots are frequencies of maximum possible insertion gain. Hence rough filter-like characteristics can be obtained by merely requiring all poles of Φ to occur in the desired attenuation bands and all roots in the desired pass bands. In order to avoid the necessity of ideal transformers or perfectly coupled inductances, however, it is best to add the restriction that Φ must be such as to make the insertion loss zero or infinite at each of the limiting frequencies zero and infinity.[68]

[68] This additional requirement is automatically satisfied when the network is to be symmetrical but not when it is to have inverse impedances. The requirement is called for in that failure to obey it always leads to the necessity of an ideal transformer or perfectly coupled inductances. On the other hand, it does not appear to be sufficient to insure that inverse impedance networks can be realized without these devices, although they are not required in the filters commonly encountered.

By way of illustration consider the special case of the low-pass filter. For a symmetrical low-pass filter

$$e^{2\alpha} = 1 + \left[S_0 \frac{\omega(\omega_1^2 - \omega^2) \cdots (\omega_\eta^2 - \omega^2)}{(1 - \nu_1^2\omega^2) \cdots (1 - \nu_\eta^2\omega^2)} \right]^2 \tag{44}$$

in which the ω_σ's, ν_σ's and S_0 are arbitrary constants. Similarly, for a low-pass filter of the inverse impedance type producing infinite loss at infinite frequency,

$$e^{2\alpha} = \frac{4R_1R_2}{(R_1 + R_2)^2} \left\{ 1 + \left[S_0 \frac{(\omega_1^2 - \omega^2) \cdots (\omega_\eta^2 - \omega^2)}{(1 - \nu_1^2\omega^2) \cdots (1 - \nu_{\eta-1}^2\omega^2)} \right]^2 \right\}. \tag{45}$$

Making each ω_σ in these expressions less than ω_c and each ν_σ less than $\frac{1}{\omega_c}$, where ω_c is the desired effective cutoff, leads to the type of low-pass filter characteristics illustrated in Figs. 17A and 17B for the case of $\eta = 3$. The insertion loss characteristics illustrated can normally be realized with the ladder networks indicated in Figs. 18A and 18B, respectively.

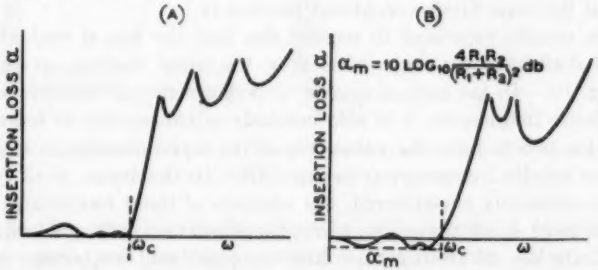


Fig. 17. Illustrations of the general form of filter characteristics obtained with symmetrical and inverse impedance networks

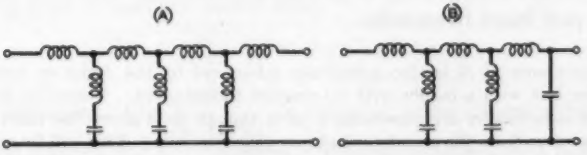


Fig. 18. Network configurations corresponding to the filter characteristics of Fig. 17

Decreasing the constant S_0 in the above expressions always decreases the pass band distortion but also decreases the suppression of attenuation band frequencies other than frequencies of infinite loss. This constant is normally made as large as is consistent with the permissible pass band distortion in order that the required suppression of unwanted frequencies may be obtained with a network of minimum complexity.[69] Zero loss at zero frequency, which is not necessarily produced by filters of the inverse impedance type, can always be obtained by properly choosing the terminations R_1 and R_2.

Tchebycheff Pass Band Parameters—Definition

Insertion loss functions of the rough filter type described above are characterized by the appearance of equal minima at arbitrary pass band frequencies and of infinite loss points at arbitrary attenuation band frequencies. Particularly useful filter characteristics and also particularly simple design procedures can be obtained by requiring the frequencies of minimum loss to be so chosen that the maxima occurring between the equal minima are themselves all equal. This leads to the type of loss characteristics illustrated in Figs. 19A and 19B for the special low-pass filters considered previously.

It is usually expedient to require also that the loss at each effective cutoff shall have the same value as at the equal maxima, as indicated in Fig. 19. In the case of inverse impedance filters transmitting zero or infinite frequencies, it is also normally advantageous to require the insertion loss to have the value zero at the equal maxima as well as at zero or infinite frequency, as in Fig. 19B. In the design of all types of filters commonly encountered, the addition of these restrictions to the requirement of equal maxima leaves the effective cutoffs and frequencies of infinite loss all arbitrary but fixes the pass band frequencies of minimum loss uniquely in terms of these parameters.[70] The characteristics obtained are the best for meeting efficiently the common type of filter requirements setting limits on permissible distortion which are constant at all pass band frequencies.

[69] The choice of S_0 is also sometimes influenced by the desire to realize the insertion loss with a ladder with no coupled inductances. Normally, however, coupled inductances are unnecessary even though S_0 is given the most advantageous value from the standpoint of loss characteristic. This will be explained later.

[70] The situation becomes more complicated only in such rare cases as multiband filters not derivable from simpler filters by means of frequency transformations.

The "equal ripple" pass band characteristics described above can be said to approximate constant losses in the Tchebycheff sense, while the corresponding pass band frequencies of minimum loss can be referred to as Tchebycheff pass band parameters. This means merely that the frequencies of minimum loss are chosen in such a way as to give the least maximum deviation from constant losses at pass band frequencies, other parameters being considered fixed.

The use of Tchebycheff parameters in network theory was first introduced by Cauer (16), who applied them to the design of image parameter filters. Filters satisfying the Tchebycheff pass band requirements of the insertion loss theory, however, are more nearly analogous to general image parameter filters than to Cauer's special Tchebycheff parameter type. A closer parallel to Cauer's use of Tchebycheff parameters is

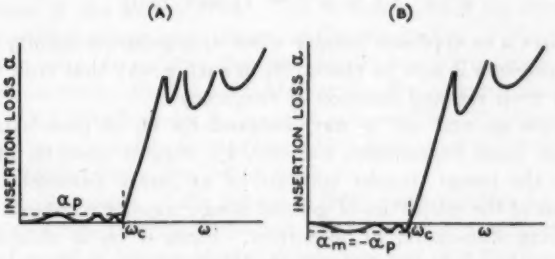


Fig. 19. Illustrations of the form of filter characteristics obtained by using Tchebycheff pass band parameters

offered by the simultaneous requirement of Tchebycheff attenuation band and pass band parameters, which will be described later.

Tchebycheff Pass Band Parameters—Theory

Tchebycheff pass band parameters can be obtained by using the following type of insertion power ratio.

$$e^{2\alpha} = \Lambda[1 + (e^{2\alpha_p} - 1)\cosh^2(\Theta_I)]. \tag{46}$$

In this expression, Λ is $\dfrac{4R_1R_2}{(R_1+R_2)^2}$ as before; α_p is the difference between the equal maxima and equal minima of the pass band insertion loss characteristic, as indicated in Fig. 19; and Θ_I is a function of frequency meeting the following requirements: First, Θ_I must be such that $\cosh(\Theta_I)$ is an odd or even rational function of frequency, depending upon

whether the power ratio is to be appropriate for symmetrical or inverse impedance filters; second, Θ_I must be pure imaginary at all pass band frequencies and must have the form $(\alpha_I + n\pi i)$ at all attenuation band frequencies, where α_I is not only real but also is infinite at various arbitrary frequencies.

It is almost obvious that the above requirements on Θ_I lead to the desired type of insertion loss characteristics. At pass band frequencies, Θ_I takes the form $i\beta_I$ leading to the power ratio

$$e^{2\alpha} = \Lambda[1 + (e^{2\alpha_p} - 1)\cos^2(\beta_I)]. \tag{47}$$

Since the value of $\cos^2(\beta_I)$ must lie between zero and unity, the corresponding value of α must lie in a range of amplitude α_p. On the other hand, at attenuation band frequencies

$$e^{2\alpha} = \Lambda[1 + (e^{2\alpha_p} - 1)\cosh^2(\alpha_I)]. \tag{48}$$

This requires α to approach infinity when α_I approaches infinity. The only real problem is how to choose Θ_I in such a way that $\cosh(\Theta_I)$ is an odd or even rational function of frequency.

The forms $i\beta_I$ and $(\alpha_I + n\pi i)$ assumed by Θ_I at pass band and attenuation band frequencies, respectively, suggest that Θ_I may be similar to the image transfer constant of an image parameter filter. An analysis of the properties of general image transfer constants shows the following statements to be correct. First, if Θ_I is obtained by adding $\frac{\pi}{2}i$ to the image transfer constant of any filter (of pure reactances) with inverse image impedances, $\cosh(\Theta_I)$ will be an odd rational function of frequency. Second, if Θ_I by itself represents an image transfer constant of any symmetrical filter, $\cosh(\Theta_I)$ will be an even rational function of frequency.[71] If the theoretical pass bands of these

[71] These relations follow directly from the following equations, which can readily be derived by means of elementary network theory:

$$\cosh\left(\Theta + i\frac{\pi}{2}\right) = i\frac{\sqrt{Z_{I1}Z_{I2}}}{-Z_{O12}}$$

$$\cosh(\Theta) = \sqrt{\frac{Z_{I1}}{Z_{I2}}}\left[\frac{-Z_{O2}}{Z_{O12}}\right]$$

in which Z_{I1} and Z_{I2} are the image impedances of a filter with transfer constant Θ while Z_{O12} and Z_{O2} are corresponding open-circuit transfer and driving-point impedances. Bode exhibits equations very similar to these. See, for instance, equation (30) on page 43 of his "General Theory of Electric Wave Filters (4)."

image transfer constants coincide with the desired effective pass bands of the insertion loss filters, Θ_I will also assume the required forms $i\beta_I$ and $(\alpha_I + n\pi i)$ at pass band and attenuation band frequencies. As a result, the well established theory of the image transfer constants of image parameter filters of the inverse impedance and symmetrical types can be applied directly to the design of insertion loss filters of the symmetrical and inverse impedance types, respectively.

Tchebycheff Pass Band Parameters—Fundamental Design Procedure

In accordance with the above principles, the design of a filter with Tchebycheff pass band parameters on the insertion loss basis is carried out in terms of the image transfer constant of a hypothetical image parameter filter. It should be borne in mind, however, that there is no direct connection between the element values of the reference filter and those of the actual filter. The reference filter involves sections with matched image impedances, or the equivalent, while the actual filter does not. In addition the reference filter is normally of somewhat greater complexity than the actual filter.[72] The reference filter is introduced at all only because its transfer constant is well understood and happens to have a functional form which facilitates the determination of the insertion loss of the actual filter.

The configuration of the reference filter corresponding to an actual filter of prescribed configuration is chosen in accordance with the requirement of equal or inverse image impedances and of a one to one correspondence between the attenuation peaks of the reference filter and all the frequencies of infinite loss of the actual filter. Since the image impedances of the reference filter are only required to be properly related, a number of configurations are always possible.

By way of illustration, the two low-pass configurations used previously as illustrative insertion loss filters are shown in Fig. 20 in comparison with the simplest corresponding reference filters. It is seen that each reference filter is obtained by adding a constant-k half-section to an image parameter filter of the same configuration as the corresponding actual filter. The added half-sections supply the infinite loss points at infinity due to the corresponding poles of the series branches of the actual filters. More generally, in the case of all types of filters commonly encountered the reference filters can be obtained by adding

[72] This is because the total insertion loss of the actual filter is determined by the transfer constant of the image parameter filter, including infinite loss points corresponding to the reflection peaks of the image parameter theory.

similar constant-k half-sections of the proper types to reference filters of the same configurations as the actual filters.[73]

When it comes to the choice of the arbitrary constants of a reference filter in such a way that the corresponding actual filter will meet design specifications of the type ordinarily encountered, the procedure is actually simpler than if the reference filter itself constituted the final network. In the first place, the theoretical pass band of the reference filter is coincident with the effective pass band of the actual filter, it being unnecessary to make any allowances for ranges of high reflection losses near theoretical cut-offs. In addition, the pass band distortion α_p can be chosen directly, rather than reflection losses due to variations in the image impedances which must be corrected for interaction effects. Finally, at attenuation band frequencies at which even moderate in-

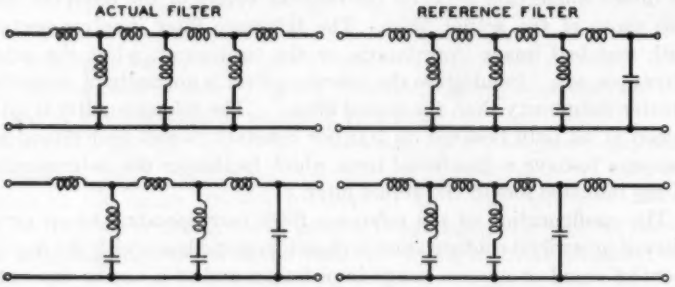


Fig. 20. Illustrations of reference image parameter filters

sertion losses are obtained, a good approximation to the loss can be computed more easily than in the case of an image parameter filter, which requires reflection losses to be added to the image attenuation. This computation is carried out by means of the following approximate formula which can be derived from equation (48) by first assuming that $\frac{1}{\Lambda}e^{2\alpha}$ and $e^{2\alpha_I}$ are both large compared with unity and then expressing α_I in terms of decibels rather than napiers:

$$\alpha - \alpha_m = [\alpha_I + 10 \log_{10} (e^{2\alpha_p} - 1) - 6.02]\ db \qquad (49)$$

where α_m is the minimum pass band insertion loss, which is negative or zero since it amounts to $10 \log_{10} (\Lambda)\ db$, while α_I is identical with the image attenuation of the reference filter.

[73] Exceptions to this general rule are possible but are not apt to be encountered in ordinary filter design problems.

After the arbitrary constants have been chosen, the function cosh (Θ_I) appearing in (46) can readily be replaced by the equivalent odd or even rational function of frequency. In the case of low-pass filters, for instance, if χ is used to represent $\sqrt{1 - \frac{\omega_c^2}{\omega^2}}$ the function cosh (Θ_I) turns out to be

$$\cosh(\Theta_I) = \frac{\prod_\sigma (m_\sigma + \chi) + \prod_\sigma (m_\sigma - \chi)}{2 \prod_\sigma \sqrt{m_\sigma^2 - \chi^2}} \tag{50}$$

In this equation, ω_c represents the value of ω at the cutoff while the m_σ's represent the "m's" determining the attenuation peaks corresponding to the various half-sections of the reference image parameter filter, there being one factor in each product corresponding to each half-section. When cosh (Θ_I) is a rational function of ω, m_σ's different from unity occur in identical pairs, corresponding to halves of m-derived full sections in the reference filter. Each such pair is related to one of the ν_σ's in the expression of the form (44) or (45) for the power ratio of the actual filter by the expression

$$\nu_\sigma = (1/\omega_c)\sqrt{1 - m_\sigma^2} \tag{51}$$

Special Formulae Aiding Root Extraction

After cosh (Θ_I) has been replaced in (46) by the equivalent rational function, corresponding networks can always be designed by means of the general theory of symmetrical and inverse impedance networks described previously. The use of Tchebycheff pass band parameters, however, permits the more general design procedure to be considerably simplified by the introduction of additional special formulae. In particular, the required root extraction can be expedited by the use of special formulae giving good approximations to the required roots. How this is accomplished is described briefly below.

The required roots are those of equations of the form

$$1 \pm i\sqrt{e^{2\alpha_\nu} - 1}\cosh(\Theta_I) = 0 \tag{52}$$

expressed as a function of $p = i\omega$.[74] In other words, the problem is to calculate the values of p corresponding to the values of Θ_I determined by

[74] These are obviously factors of the function $\frac{1}{A} e^{2\alpha}$ determined by (46). In the case of symmetrical filters only one choice of sign need be considered, the

$$\Theta_I = \cosh^{-1}\left[\frac{\pm i}{\sqrt{e^{2\alpha_p} - 1}}\right] \tag{53}$$

If the reference image parameter filter is made up of identical m-derived sections or half-sections, Θ_I will take the form $(n\Theta_0)$ or $\left(n\Theta_0 + \frac{\pi}{2}i\right)$, in which Θ_0 represents the transfer constant of a single section or half-section. When the reference filter is not actually made up of identical sections, Θ_I can normally be approximated over a portion of the p plane including the roots of (52) by means of the Θ_I function representing a reference filter of this special type. In ordinary design problems it is only necessary to replace the "m's" describing the various sections of the reference filter with a single new "m" representing their mean.[75] Since p can normally be expressed as a simple function of Θ_0, it follows that approximations to the roots of (52) in terms of p can ordinarily be computed by replacing Θ_I by $(n\Theta_0)$ or $\left(n\Theta_0 + \frac{\pi}{2}i\right)$ in (53). When Θ_0 is the transfer constant of an m-derived low-pass half-section of "m" equal to m_0, for instance, the approximations to the roots of (52) can be calculated by means of the following formula in terms of the values of Θ_0 determined by (53):

$$p = \frac{\omega_c \sinh(\Theta_0)}{\sqrt{1 - (1 - m_0^2)\cosh^2(\Theta_0)}} \tag{54}$$

in which the sign of the square root must be such that its real part is positive.[76]

It can readily be shown that the approximations to the roots described above can be expected to be reasonably good in ordinary filter design

---

corresponding roots being those of $(P + pB')$, which are required in forming (41). In the case of inverse impedance filters the roots corresponding to one choice of signs are the conjugates of those corresponding to the other.

[75] The approximating Θ_I function does not have to represent a symmetrical or inverse impedance filter as does the actual Θ_I function. In the case of low-pass or high-pass filters, the harmonic mean of the "m's" appears to have particular advantages. The same thing is true of band pass or band elimination filters if they have symmetrical loss characteristics permitting the m's to represent the derivation of confluent sections from constant-k sections.

[76] Pure imaginary values of the square root, for which the sign is not defined, are not encountered in problems in which (54) gives reasonable approximations to the required roots.

problems. Experience has shown that they are normally good enough for well-known root improvement methods to be applied directly to the determination of the actual roots to any desired precision. It also turns out that the approximate roots are of such a form that special methods can be used to obtain second approximations before the application of general root improvement methods.[77]

Other Special Formulae

Various other special formulae can be derived to facilitate the design of filters on the Tchebycheff pass band parameter basis besides those expediting the root extraction. The reflection coefficients, for instance, which measure the departure of the driving-point impedances from the corresponding terminating impedances satisfy the following limit at pass band frequencies:

$$\left|\frac{R-Z}{R+Z}\right| \leqq \sqrt{1-e^{-2\alpha_p}} \tag{55}$$

in which Z is either driving-point impedance and R is the corresponding termination.

A rough estimate of the variation α_d of the insertion loss over the pass band which will be produced by parasitic dissipation in a low-pass or high-pass filter is usually furnished by the highly approximate relation,[78]

$$e^{\alpha_d} = 1 + \frac{2n}{Qm_0^2} \coth\left[\frac{1}{n}\coth^{-1}(e^{\alpha_p})\right] \tag{56}$$

In this expression Q is the harmonic mean of the magnitudes of the reactance-resistance ratios of all the inductances and capacities, evaluated at the cut-off frequency; m_0 represents the harmonic mean of the "m's" of all the half-sections in the reference image parameter filter, and n is the total number of half sections in the reference filter.

[77] This is accomplished by using a simple transformation of variable which transforms the approximate roots into the n-th roots of a constant. One method of obtaining second approximations amounts to the use of Newton's method in conjunction with Fry's isograph. Although this machine is primarily intended for general root extraction purposes it is also convenient for applying Newton's method to the improvement of roots approximated by quantities which are all of the same magnitude.

[78] This equation furnishes even rough estimates only in the case of filters meeting ordinary specifications, particularly as regards sharpness of cutoffs and non-dissipative pass band loss.

The actual loss α_∞ produced by a low-pass or high-pass ladder filter at a frequency of infinite theoretical loss not too far removed from the cutoff frequency can be estimated by means of the following formula:

$$e^{\alpha_\infty} = \frac{2m_\infty^2 Q}{(1 - m_\infty^2)}\, e^{\alpha'} \tag{57}$$

In this equation, m_∞ is the "m" corresponding to the particular peak frequency considered and α' is calculated by subtracting the attenuation of the corresponding section of the reference filter from the non-dissipative insertion loss of the actual filter and evaluating the difference at the peak frequency. Q is now best chosen as the mean "Q" of the particular elements of the ladder network whose resonance produces the peak, the ratios being evaluated at the peak frequency.

Ordinarily, equation (56) can be applied to band-pass and band-elimination filters with narrow pass or attenuation bands by merely multiplying Q by $\frac{f_{c2} - f_{c1}}{2\sqrt{f_{c2}f_{c1}}}$, where f_{c2} and f_{c1} are the cutoff frequencies and Q is now evaluated at the mean frequency of the pass or attenuation band. This assumes, however, that the loss characteristics are at last approximately symmetrical on a logarithmic frequency scale. Equation (57) can also be applied to these filters provided the peak frequencies considered are not too far removed from cutoff frequencies. This application is accomplished by multiplying Q by $\frac{|f_\infty^2 - f_{c2}f_{c1}|}{2f_\infty\sqrt{f_{c2}f_{c1}}}$, in which f_∞ is the peak frequency.

Comparison of Actual Performance of Tchebycheff Pass Band Parameter Filters with That of Image Parameter Filters

From the standpoint of actual performance, a Tchebycheff pass band parameter filter can best be compared with an image parameter filter producing the same order of magnitude of distortion at frequencies in the effective pass band. When the constant α_p measuring the pass band distortion of the insertion loss filter is of the order of tenths of a decibel, for instance, the best comparison is usually obtained by assuming the image parameter filter to have image impedances of the constant-k type. When α_p is of the order of hundredths of a decibel or smaller, on the other hand, it is usually best to assume that the reflection effects produced by the image parameter filter at pass band frequencies are reduced by the use of image impedances including impedance controlling factors.

The comparison is most clear cut when α_p is of the order of the distortion due to an image parameter filter with constant-k image impedances. Recall that the reference filter used in designing an insertion loss filter is usually obtained by adding a constant-k half-section to an image parameter filter of the same configuration as the insertion loss filter. In other words, equation (49) approximating the insertion loss of an actual filter at frequencies of high loss can usually be written in the form

$$\alpha - \alpha_m = \alpha_I' + [\alpha_0 + 10 \log_{10} (e^{2\alpha_p} - 1) - 6.02]\ db \qquad (58)$$

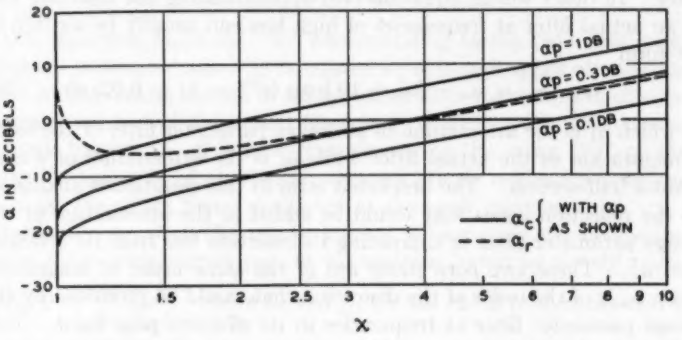
in which α_I' is the attenuation of an image parameter filter of the same configuration as the actual filter while α_0 is the attenuation of a constant-k half-section. The bracketed term in this equation is analogous to the reflection losses that would be added to the attenuation of the image parameter filter in computing its insertion loss from its attenuation α_I'. These two corrections are of the same order of magnitude when α_p is of the order of the distortion that would be produced by the image parameter filter at frequencies in its effective pass band. This is illustrated by actual curves of the two corrections in Fig. 21, assuming low-pass filters.

Although the insertion loss filter described above and the corresponding image parameter filter of the same configuration produce insertion losses of the same order of magnitude at frequencies of high loss, their pass band loss characteristics differ in one important respect. The effective cutoff of the insertion loss filter coincides with the theoretical cutoff of the image parameter filter, while the effective cutoff of the image parameter filter normally occurs at a frequency at least 10 per cent lower.[79] This means that if the image parameter filter were to be redesigned so as to have the same effective cutoff as the insertion loss filter the theoretical cutoff would have to be moved much nearer to the attenuation peaks. If the theoretical cutoff were originally within 20 per cent of an attenuation peak, for instance, it would normally have to be moved to within 10 per cent of the peak. This change would produce a marked reduction in the attenuation band loss, as is indicated by a glance at a set of attenuation curves for filters of different peak frequencies, such as that exhibited by Shea (17).

As is indicated by the above discussion, insertion loss filters of the

[79] The effective cutoff of the image parameter filter can of course be moved closer to the theoretical cutoff if impedance controlling factors are used but this requires an increase in network complexity.

Tchebycheff pass band parameter type normally represent a substantially more efficient use of the elements than image parameter filters when the pass band distortion is to be of the order of tenths of a decibel. The same general situation usually exists when the pass band distortion is to be substantially lower, calling for an image parameter filter in-



$$\alpha_c = \alpha_s + 10 \log_{10} (e^{2\alpha_r} - 1) - 6.02 \text{ db}$$

$$\alpha_s = \tanh^{-1} \sqrt{1 - \frac{1}{x^2}}$$

= attenuation of a constant-k low-pass half-section.

$$\alpha_r = 20 \log_{10} \left(\frac{x^2}{4\sqrt{x^2 - 1}} \right)$$ = reflection loss between two constant-k low-pass image impedances and their nominal impedances.

$$x = \frac{\omega}{\omega_c}$$

Fig. 21. A comparison of the corrections involved in the determination of the insertion losses of low-pass filters of the Tchebycheff pass band parameter and image parameter types

volving impedance controlling factors, but not necessarily to the same extent.

Dependence Upon α_p of Positiveness of Ladder Elements

In Part II certain conditions were pointed out which are sufficient to insure that a mid-series low-pass ladder will not include negative inductances requiring the use of coupled coils. The first of these requires the frequencies of infinite loss to be greater than all the finite poles of at least one of the open-circuit driving-point impedances. The

second requires the particular frequencies of infinite loss corresponding to the shunt branches nearest the terminals to be also equal to or greater than all the roots of the corresponding open-circuit impedances. As would be expected, the open-circuit impedances of a Tchebycheff pass band parameter filter vary in much the same way as those of a corresponding image parameter filter producing a somewhat similar loss characteristic. Specifically, a good comparison is usually obtained when the two filters produce the same order of magnitude of pass band distortion and are described by impedances which vary in the same manners at zero and infinite frequencies, rather than in inverse manners. Thus a study of the impedances of image parameter filters gives an indication of when coupled coils may be required in insertion loss filters.

The open-circuit impedances of an image parameter filter are given by

$$Z_0 = Z_I \coth (\Theta) \tag{59}$$

where Z_0 is either open-circuit impedance, Z_I is the corresponding image impedance, and Θ is the image transfer constant. Since coth (Θ) is finite in attenuation bands, the only roots and poles of the open-circuit impedances occurring in attenuation bands are roots and poles of the image impedances.

A mid-series low-pass image impedance of the constant-k type has no root above the cutoff frequency and has no pole except at infinity. It follows that low-pass filters of the insertion loss type and filters related to them by frequency transformations can be expected to be realizable with ladders without coupled coils whenever the pass band distortion is of the order of that produced by image parameter filters with constant-k impedances.

When a single impedance controlling factor is considered, a mid-series type of low-pass insertion loss filter must be compared with an image parameter filter of the so-called mid-shunt m-terminated type, which produces an image impedance with a pole at infinity.[80] In this case,

[80] The well known ladder form of a symmetrical mid-shunt m-terminated image parameter filter has the general mid-series low-pass configuration except that the terminal series inductances do not appear. The insertion power ratio of a general mid-series low-pass ladder, when expressed in the form (35), has as many constants as there are elements in the network. It turns out that removing the terminal series branches leaves the form and degree of the power ratio unchanged, the only result being changes in the numerical values of the constants, which are greater in number than the elements of the "reduced" network. Thus the mid-shunt m-terminated image parameter low-pass filter can be described as a special case of the general mid-series low-pass configuration in which the terminal series inductances have the value zero.

the open-circuit impedances of the image parameter filter will have roots in the attenuation band but no poles except at infinity. Thus when an insertion loss filter produces a pass band distortion comparable with that obtained with image impedances each including a single impedance controlling factor, it is to be expected that no coupled coils will be required provided the frequencies of infinite loss produced by the terminal shunt branches are sufficiently high.

When two or more impedance controlling factors are included in each image impedance, there will be poles in attenuation bands and the sufficient conditions for elimination of coupling will be violated unless all infinite loss frequencies are sufficiently high. This does not mean, however, that coupling will necessarily be required whenever there are infinite loss frequencies near the cutoff, since the above conditions are sufficient but not necessary for its elimination. It merely means that the possibility of required coupling must be considered.

The general rule is obviously that reducing the pass band distortion α_p always makes the necessity of coupled coils more probable. For the range of values of α_p normally of interest it can usually be assumed that no coupling will be necessary, but unusually small values of α_p may lead to its necessity.[81]

Simultaneous Attenuation Band and Pass Band Tchebycheff Parameters

In many filter design problems, the minimum permissible suppression of unwanted frequencies is constant over prescribed effective attenuation bands at the same time that the maximum permissible distortion is constant over prescribed effective pass bands. In many such problems, the most efficient use of the arbitrary constants is obtained by requiring Tchebycheff attenuation band parameters as well as Tchebycheff pass band parameters. In other words, the frequencies of infinite loss of the previous theory are required to be so chosen that the insertion loss characteristics exhibit equal minima between the frequencies of infinite loss as well as equal maxima and equal minima in the pass bands. The simplest illustration is the design of a low-pass filter with an effective pass band extending from zero frequency to some prescribed frequency f_1 and an effective attenuation band extending

[81] The values of α_p leading to the requirement of coupling are normally extremely small, but very small values are sometimes required in order to meet restrictions on permissible impedance variations. As is indicated by (55), the impedance variations occurring at pass band frequencies approach zero only as $\sqrt{\alpha_p}$.

from a second prescribed frequency f_2 to infinity.. In this special case, the simultaneous use of attenuation band and pass band Tchebycheff parameters described above leads to a loss characteristic of the type illustrated in Fig. 22 for a three-section symmetrical filter.

Mathematically, the theory of simultaneous attenuation band and pass band Tchebycheff parameters is very similar to the Tchebycheff version of the image parameter theory introduced by Cauer (16). Design formulae for the essential constants such as the frequencies of infinite loss can actually be derived by combining Cauer's theory with the theory of general Tchebycheff pass band parameters described above. Cauer's theory leads to image parameter filters characterized by equal minima of attenuation between attenuation peaks. If one of these is used as the reference image parameter filter in the design of an insertion

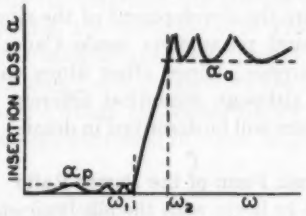


Fig. 22. An illustration of the form of filter characteristics obtained by the use of both attenuation band and pass-band Tchebycheff parameters

loss filter with Tchebycheff pass band parameters, equal minima of insertion loss between the infinite loss points are obtained. The requirement that cosh (Θ_I) must be an even or odd rational function of frequency and must be infinite at a zero or infinite attenuation band frequency, however, requires Cauer's basic theory to be modified somewhat in designing the image parameter filter.[52]

The similarity of the insertion loss filters to Cauer's image parameter filters does not extend beyond the theory involved in the choice of the design parameters. The insertion loss characteristics produced by the two types of filters are quite different. Cauer's theory would lead to insertion loss characteristics with "equal ripples" in pass bands and with equal minima between infinite loss points in attenuation bands only if

[52] The required modification can be determined by examining Cauer's various methods of modifying his theory to apply it to various image parameter problems.

interaction effects could be neglected at pass band frequencies and reflection effects at attenuation band frequencies. In actual filters, the effects of interaction and reflections are sufficient to produce marked changes in the loss characteristics.[83] The insertion loss filters, on the other hand, produce the ideal characteristics exactly, there being no corrections for such effects as reflections and interaction.[84]

Direct Derivation of the Theory

In spite of the fact that Cauer's Tchebycheff version of the image parameter theory can be used to derive the essential formulae, it is more convenient to develop the corresponding insertion loss theory by means of the more direct analysis described below. This analysis yields the same formulae as Cauer's theory and also additional useful relationships. It was used in the original derivation of the theory, which was carried out before the development of the simple theory of general Tchebycheff pass band parameters made Cauer's image parameters available for the purpose. Since other filters call for a very similar method of analysis although somewhat different formulae, only symmetrical low-pass filters will be described in detail.

Statement of Algebraic Form of the Power Ratio

It will be simplest to begin with the algebraic statement of the insertion power ratio required for simultaneous attenuation band and pass band Tchebycheff parameters, leaving its justification until later. An expression for the power ratio exhibiting a simple type of symmetry is obtained by replacing ω by a new variable Ω even though it is a low-pass filter which is under consideration. In the design of a low-pass filter, Ω is defined by

$$\Omega = \frac{\omega}{\sqrt{\omega_1 \omega_2}} \tag{60}$$

[83] At a pass band frequency at which the phase is an integral multiple of π radians, for instance, interaction effects in a symmetrical image parameter filter are such as to produce zero insertion loss whatever the reflection losses. At attenuation band frequencies the reflection effects vary from 6db reflection gain to infinite reflection loss.

[84] Even when dissipation in the elements is considered, it is frequently possible to modify the insertion loss filters in such a way that the insertion loss characteristics differ from those of non-dissipative filters only by constant losses. How this can be done will be explained in Part IV.

where ω_1 and ω_2 are the values of ω at the cutoff frequency f_1 and the limit of the effective attenuation band f_2. The values of Ω at f_1 and f_2 are then $\sqrt{k}$ and $\frac{1}{\sqrt{k}}$ where

$$k = \frac{f_1}{f_2} \tag{61}$$

The corresponding power ratio required for a symmetrical filter turns out to be of the form

$$e^{2\alpha} = 1 + \left[S_0 \frac{\Omega(\Omega^2 - \Omega_1^2) \cdots (\Omega^2 - \Omega_\eta^2)}{(1 - \Omega_1^2\Omega^2) \cdots (1 - \Omega_\eta^2\Omega^2)} \right]^2 \tag{62}$$

Consider the bracketed expression in (62). If the constant S_0 is omitted, replacing Ω by its reciprocal will replace this expression by its reciprocal. It follows that the requirement of Tchebycheff attenuation band parameters will be satisfied provided the Ω_σ's are chosen in such a way that Tchebycheff pass band parameters are obtained. As would be expected from Cauer's image parameter theory, the required values of the Ω_σ's are most simply expressed in terms of Jacobean elliptic functions. Specifically

$$\Omega_\sigma = \sqrt{k}\, \mathrm{sn}\left[\frac{2\sigma K}{2\eta + 1}, k \right] \qquad \sigma = 1, \cdots, \eta \tag{63}$$

in which the constant k representing the modulus still has the value f_1/f_2, while K is the corresponding complete integral.

Proof of Correctness of the Algebraic Expression for the Power Ratio

Suppose the power ratio (62) is transformed by replacing Ω by $\sqrt{k}\, \mathrm{sn}\,(u, k)$. The function of u obtained in this way can be simplified by replacing the Ω_σ's by their equivalents of (63) and then making use of the summation law for elliptic sines. This procedure permits the bracketed expression in (62) to be replaced by a product of elliptic sines showing the power ratio to be determined by the following pair of equations

$$e^{2\alpha} = 1 + S_0^2 \prod_{\sigma=-\eta}^{\sigma=+\eta} \left\{ k\, \mathrm{sn}^2 \left[\left(u + \frac{2\sigma K}{2\eta + 1} \right), k \right] \right\} \tag{64}$$

$$\Omega = \sqrt{k}\, \mathrm{sn}\,(u, k).$$

If u is replaced by $\left(u + \frac{2K}{2\eta + 1}\right)$ in the product in the above power ratio the net effect is to replace the factor in $\sigma = -\eta$ by one in $\sigma = \eta + 1$, the other factors being merely interchanged. The arguments of the factors in $\sigma = -\eta$ and $\sigma = \eta + 1$ differ by exactly $2K$ and must therefore be identical since this is the period of $\mathrm{sn}^2(u, k)$. In other words, the power ratio is unchanged by replacing u by $\left(u + \frac{2K}{2\eta + 1}\right)$, which amounts to saying it is periodic in u with the period $\frac{2K}{2\eta + 1}$. The frequency variable Ω is also periodic in u but has the longer period $4K$.

For real values of u the function $\mathrm{sn}(u, k)$ covers the range -1 to $+1$. In other words, the corresponding values of Ω coincide with the pass band and its image at negative frequencies. Thus as u increases Ω will vary cyclically over the pass band and its negative image, while in each cycle of Ω the power ratio will pass through several complete cycles, i.e., through a series of equal maxima and equal minima. A more detailed analysis of this situation shows that it requires the power ratio under consideration to represent at least a special case of Tchebycheff pass band parameters.[85] As indicated above, it then follows from the reciprocal nature of the bracketed expression in (62) that Tchebycheff attenuation band parameters are also obtained.[86]

Determination of the Roots of the Power Ratio and of Related Functions

The periodicity of the power ratio and of Ω considered as functions of u simplifies the calculation of the roots of the power ratio involved in the determination of corresponding networks. Since the period of the power ratio is a fraction of that of Ω, as soon as any one root of the

[85] The general principle of using periodic transformations in Tchebycheff parameter problems was introduced by Schelkunoff. Schelkunoff's application of the principle to Cauer's Tchebycheff theory is described by Guillemin (18).

[86] The above analysis merely proves the correctness of a particular solution for the choice of the Ω_∞'s yielding attenuation band and pass band Tchebycheff parameters. The uniqueness of this solution can be demonstrated, however, by making use of the following relation, which can be shown to be necessarily satisfied:

$$\frac{\pm dy}{\sqrt{(k_1 - y^2)(1 - k_1 y^2)}} = (2\eta + 1)\frac{K_1}{K}\frac{d\Omega}{\sqrt{(k - \Omega^2)(1 - k\Omega^2)}}$$

in which y represents the bracketed expression in (62) with the constant S_0 omitted.

power ratio in terms of $i\Omega$ has been determined a number of others can be computed by merely adding multiples of the period of the power ratio to the corresponding value of u. Actually, it turns out that all the roots can be found in this way as soon as any one has been determined. If $\pm a_0$ is used to represent the only real roots, for instance, the complex roots take the form

$$i\Omega_\sigma = \frac{\pm a_0 cd_\sigma \pm iW\Omega_\sigma}{1 + a_0^2\Omega_\sigma^2} \qquad \sigma = 1, \cdots, \eta \quad (65)$$

where Ω_σ is again $\sqrt{k}\,\mathrm{sn}\left[\frac{2\sigma K}{2\eta+1}, k\right]$ while W and cd_σ are defined by

$$W = \sqrt{(1 + ka_0^2)\left(1 + \frac{1}{k}a_0^2\right)}$$
$$cd_\sigma = \mathrm{cn}\left[\frac{2\sigma K}{2\eta+1}, k\right]\cdot\mathrm{dn}\left[\frac{2\sigma K}{2\eta+1}, k\right] \quad (66)$$

Assuming a_0 to be the *positive* real root of the power ratio, the roots of the related polynomial $(A + pB)$ appearing in the general theory of Part I are found by using $-a_0$ as the real root and replacing $\pm a_0$ by $-a_0$ in the above formulae for the complex roots.[87] It also turns out that the roots of $(P + pB')$, which are required in the formation of equation (41) in the determination of corresponding lattices, are obtained by using $+a_0$ as the real root and replacing $\pm a_0$ by $(-)^\sigma a_0$ in the above formulae.

Introduction of a Modular Transformation on the Elliptic Functions

Formulae for computing the real zeros $\pm a_0$ and other extremely useful design relations can be obtained by the introduction of a modular transformation on the elliptic functions.[88] By means of this trans-

[87] These formulae of course give the roots in terms of $i\Omega$, defined as $\frac{i\omega}{\sqrt{\omega_1\omega_2}}$, rather than in terms of $i\omega$. The transformation from one variable to the other is carried out at any convenient point in the design procedure, commonly after the impedances of the final network branches have been determined as functions of $i\Omega$, as in the case of more complicated frequency transformations.

[88] The general theory of modular transformations (other than Landen's transformation and Jacobi's imaginary transformation) is described in very few treatises on elliptic functions. It is described in detail by Cayley (19), however, and also by Jacobi (20), who was the first to introduce it. The use of modular transformations in connection with the insertion loss theory was introduced by Norton, who also showed that their use simplifies Cauer's Tchebycheff version of the image parameter theory.

formation, the product of elliptic functions appearing in the power ratio expression (64) can be replaced by a single elliptic function of a different modulus. This transformation yields an expression of the form

$$e^{2\alpha} = 1 + (e^{2\alpha_p} - 1)\,\mathrm{sn}^2\left[(2\eta + 1)u\frac{K_1}{K}, k_1\right] \tag{67}$$
$$\Omega = \sqrt{k}\,\mathrm{sn}\,(u, k)$$

where K_1 represents the complete integral of the new modulus k_1 while α_p is again the magnitude of the equal pass band ripples of the insertion loss characteristic. The moduli k_1 and k are uniquely related in the manner described below.

Suppose K' and K_1' represent the complete elliptic integrals of moduli $\sqrt{1-k^2}$ and $\sqrt{1-k_1^2}$, respectively, just as K and K_1 designate the complete integrals of moduli k and k_1. Then the so-called modular constants q and q_1 are defined as $\exp\left(-\pi\frac{K'}{K}\right)$ and $\exp\left(-\pi\frac{K_1'}{K_1}\right)$, respectively. They are uniquely related to k and k_1 and are required to satisfy the equation

$$q_1 = q^{2\eta+1}. \tag{68}$$

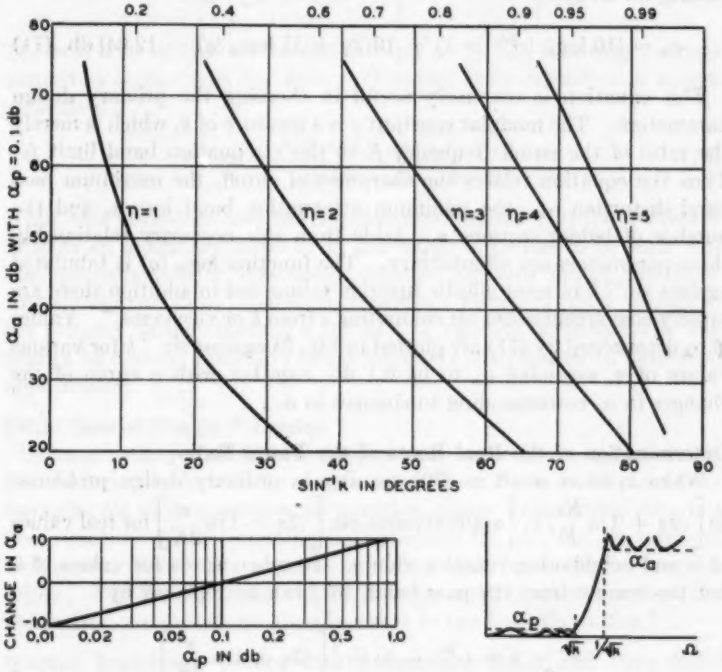
The above relation depends upon the fact that the equivalent power ratios (64) and (67) must both have the periods $\frac{2K}{2\eta+1}$ and $2iK'$ in terms of u. The equivalence of the power ratios can be established by showing that when the periods are the same the roots, singularities, and behavior at infinity of the functions $(e^{2\alpha} - 1)$ are identical, which requires them to be proportional in accordance with Liouville's theorem. The constant or proportionality, which relates S_0 of (64) to α_p of (67), can be found by ordinary elliptic function analysis.

Relation between Arbitrary Design Parameters

One useful formula that can be derived from (67) is a very simple relation between the arbitrary design parameters. In the first place, well-known elliptic function relations combined with the identity of q_1 and $q^{2\eta+1}$ show (67) to be equivalent to

$$e^{2\alpha} = 1 + \frac{e^{2\alpha_p} - 1}{k_1^2\,\mathrm{sn}^2\left[(2\eta+1)u_A\frac{K_1}{K}, k_1\right]} \tag{69}$$
$$\Omega = \frac{1}{\sqrt{k}\,\mathrm{sn}\,(u_A, k)}$$
$$u_A = u + iK'.$$

For real values of u_k, the elliptic sines in these equations will vary between $+1$ and -1. This requires the corresponding values of Ω to coincide with the effective attenuation band and its image at negative



$\alpha_a = [10 \log_{10} (e^{2\alpha_p} - 1) - 10(2\eta + 1) \log_{10} q - 12.04]$ db
q = a function of k
η = number of ladder sections

Fig. 23. Tchebycheff attenuation band and pass band parameters—approximate relation between the design parameters

frequencies. It then follows that the loss α_a at the equal minima of the effective attenuation band will be such that

$$e^{2\alpha_a} = 1 + \frac{e^{2\alpha_p} - 1}{k_1^2} \tag{70}$$

It is obvious from this equation that k_1^2 will be quite small in the case of ordinary filters, for which α_p is small and α_a large. When k_1^2

is small, moreover, the corresponding modular constant q_1 approximates $\frac{1}{16}k_1^2$ very closely. Introducing this approximation and the identity of q_1 and $q^{2\eta+1}$ in (70), assuming $e^{2\alpha_a}$ large compared with unity, and solving for α_a gives

$$\alpha_a = [10 \log_{10} (e^{2\alpha_p} - 1) - 10(2\eta + 1) \log_{10} (q) - 12.04] \text{ db} \quad (71)$$

This equation is extremely useful in choosing the primary design parameters. The modular constant q is a measure of k, which is merely the ratio of the cutoff frequency f_1 to the attenuation band limit f_2. Thus the equation relates the sharpness of cutoff, the maximum pass band distortion α_p, the minimum attenuation band loss α_a and the number of ladder sections η. Aside from this necessary relationship these parameters are all arbitrary. The function $\log_{10} (q)$ is tabulated against $\sin^{-1} k$ in most elliptic function tables and in addition there are rapidly convergent series for computing q from k or vice versa.[39] Values of α_a determined by (71) are plotted in Fig. 23 against $\sin^{-1} k$ for various values of η, assuming α_p to be 0.1 db, together with a curve of the changes in α_a corresponding to changes in α_p.

Determination of the Real Roots of the Power Ratio

When k_1 is as small as (70) requires in ordinary design problems, $\text{sn}\left[(2\eta + 1)u\frac{K_1}{K}, k_1\right]$ approximates $\sin\left[(2\eta + 1)u\frac{\pi}{2K}\right]$ for real values of u and neighboring complex values. In other words for values of Ω not too remote from the pass band, (67) can be replaced by

$$\begin{aligned} e^{2\alpha} &= 1 + (e^{2\alpha_p} - 1) \sin^2\left[(2\eta + 1)u\frac{\pi}{2K}\right] \\ \Omega &= \sqrt{k}\ \text{sn}\ (u, k). \end{aligned} \quad (72)$$

The function sn (u, k) cannot normally be replaced by $\sin\left[u\frac{\pi}{2K}\right]$ since the modulus k represents the ratio of the cutoff frequency to the attenuation band limit and is rarely small.

The above equations can be used to compute the real roots $\pm a_0$ of the power ratio, which must be determined before the formulae (65)

[39] For a tabulation of $\log_{10} (q)$ see, for instance, Silberstein's "Synopsis of Applicable Mathematics" (21).

for the complex zeros can be used. If u is replaced by iu', the above power ratio will be zero at the real value of u' determined by

$$\sinh\left[(2\eta+1)u'\frac{\pi}{2K}\right]=\frac{1}{\sqrt{e^{2\alpha_p}-1}} \tag{73}$$

The real zero a_0 can be calculated from u' by the following formula derived by replacing sn (iu', k) in $i\sqrt{k}$ sn (iu', k) by Jacobi's well known equivalent:

$$a_0=\frac{\sqrt{k}\,\mathrm{sn}\,(u', \sqrt{1-k^2})}{\mathrm{cn}\,(u', \sqrt{1-k^2})} \tag{74}$$

While this procedure uses the approximate formula (72) for the power ratio at an imaginary value of u, this value turns out to be such that the approximation is reasonably close in ordinary design problems. Any small error in a_0, moreover, represents merely a small change in the primary design constant α_p or α_a. A rigorous formula for u' can readily be derived from (67) but is relatively complicated in numerical applications.

Other Special Design Formulae

A large variety of additional rigorous and approximate special design formulae can be derived by similar methods. These include not only formulae for such quantities as insertion phase, impedances, etc., but also equations for determining the constants Ω_σ, a_0, etc., in terms of well-known Θ function expansions rather than from elliptic function tables. For the particular elliptic function computations required the Θ function expansions are usually easier to use than the tables.[90]

Inverse Impedance Filters with Attenuation Band and Pass Band Tchebycheff Parameters

Inverse impedance filters with attenuation band and pass band Tchebycheff parameters call for exactly the same sort of analysis as symmetrical filters. The only real difference is in the specific periodic substitution and elliptic function transformation involved. The simplest formulae are obtained by requiring zero loss at zero frequency and at the pass band minima rather than at zero frequency, at the cutoff,

[90] These Θ function expansions are the Fourier series representations of the functions. They are particularly convenient in that the arguments of the trigonometric functions involved depend only on η. The expansions are listed in various tables of formulae such as Silberstein's "Synopsis of Applicable Mathematics" (21).

and at the pass band maxima. While this leads to a somewhat less efficient use of the elements it at least permits the terminations to be equal. The corresponding power ratio turns out to be determined by the following set of equations

$$\begin{aligned} e^{2\alpha} &= 1 + (e^{2\alpha_p} - 1)\,\mathrm{sn}^2\left[(2\eta)u\,\frac{K_1}{K}, k_1\right] \\ \Omega^2 &= k\,\mathrm{sn}\,[u, k]\,\mathrm{sn}\left[\left(u + \frac{K}{\eta}\right), k\right] \\ q_1 &= q^{2\eta} \\ f_1/f_2 &= k\,\mathrm{sn}^2\left[\frac{2\eta - 1}{2\eta}\,K, k\right] \end{aligned} \tag{75}$$

In these equations, Ω and f_1, f_2 are defined as in the case of symmetrical filters but k no longer represents f_1/f_2, being now defined by the last of these equations.

The more efficient requirement of zero loss at zero frequency and at the pass band maxima can be satisfied by making linear transformations on $e^{2\alpha}$ and Ω^2 in these equations. The increased efficiency is indicated by the corresponding change in the formula for f_1/f_2, which becomes

$$f_1/f_2 = k\,\mathrm{sn}\left[\frac{2\eta - 1}{2\eta}\,K, k\right] \tag{76}$$

Other Types of Filters

Other types of filters with attenuation band and pass band Tchebycheff parameters besides the low-pass types described above can be obtained by means of ordinary frequency transformations. These include high-pass filters and also band-pass and band-elimination filters with symmetrical loss characteristics on a logarithmic frequency scale. More general types of characteristics can be obtained by making rational transformations on the power ratio and on the square of the frequency, but this usually leads to power ratios not realizable with networks of the symmetrical or inverse impedence types. Still more general characteristics can be obtained by combining similar transformations with the use of Tchebycheff pass band parameters but general attenuation band parameters.

Instead of transforming the power ratios described previously, the analysis used in driving them can itself be modified in such a way as to

lead to new types of loss characteristics. Differences between the transfer constants of reference image parameter filters with matched impedances can be used, for instance, in the design of filters on the Tchebycheff pass band parameter basis.[91] Similarly, new types of loss characteristics can be obtained by using different elliptic functions in (67). Tchebycheff attenuation band parameters can be obtained with general pass band parameters by merely replacing cosh (Θ_I) by sech (Θ_I) in (46).

Even when the power ratio is appropriate for symmetrical or inverse impedance filters, the terminations can both be chosen arbitrarily if the requirement of networks of these particular types is abandoned.[92] It is always possible, for instance, to make one termination zero or infinite.

Part IV. Dissipative Reactance Networks

The parasitic dissipation which must be present in actual reactance networks is referred to in Parts I, II and III only in connection with methods of estimating the effects of the dissipation in networks designed on a non-dissipative basis. These effects, however, frequently represent substantial differences between the characteristics actually realized and those which the non-dissipative design procedure assumes to be obtained. This makes it desirable to include in the insertion loss theory a modification applying directly to the design of dissipative reactance networks.

The logical procedure is to parallel the theory of non-dissipative networks with a synthesis theory applying to the design of networks which are made up of dissipative reactance elements and which produce prescribed insertion loss characteristics when terminated in prescribed resistances. The filter theory of Part III can then be applied to the choice of loss characteristics provided this can be done without violating the conditions necessary for the realizability of dissipative networks of the type required. Even when the configurations are required to be those which would be used in the absence of dissipation there frequently turn out to be realizable loss functions at least very closely approxi-

[91] This amounts to saying that sections described by negative "m's" can be included in the reference filters.

[92] Provided, of course, $\dfrac{4R_1R_2}{(R_1+R_2)^2}$ is equal to or less than the smallest value assumed by $e^{2\alpha}$ at real frequencies.

mating the desired special functions of Part III except for added constant losses.[93]

The best method to use in designing a dissipative reactance network depends upon what distribution of the dissipation is required—i.e., upon the relative dissipativeness of the various elements to be used. In the networks encountered in ordinary design problems the dissipation approximates closely certain very simple distributions for which special design procedures can be developed. This situation makes it unnecessary to develop a perfectly general theory permitting the dissipation constants of all the elements to be prescribed independently. Any such general theory must obviously be undesirably complicated since it must permit the introduction of a large number of additional arbitrary constants besides those determining the insertion loss function.

Equally Dissipative Inductances and Capacities

The simplest theory is obtained by assuming a dissipative reactance network meeting the following very special requirements. Each actual inductance is required to be equivalent to an ideal inductance combined with a constant series resistance. Similarly, each actual capacity must be equivalent to an ideal capacity combined with a constant shunt resistance. Finally, all the inductance-resistance and capacity-conductance ratios are required to be equal, which amounts to requiring the phase angles of the impedances of all the reactance elements to be of equal magnitude.

Fundamental Transformation Principle

To understand the design of networks involving the uniform type of dissipation described above it is simplest to begin by examining the effect of adding such dissipation to all the reactance elements in any known network. The determination of the function of p representing any voltage ratio, impedance, or similar complex quantity describing the known network is accomplished by properly combining the im-

[93] Bode (22) discovered ways of adding resistances to image parameter filters in such a manner as to reduce the distortion due to dissipation. This procedure increases the number of reactance elements, however, by requiring certain elements to be split in two, and in addition it only partially compensates for the effects of dissipation. Bode also discovered certain general types of equalizers which can actually duplicate the discrimination characteristics of filters and in which all inductances can be in series with resistances and all capacities in parallel with resistances and can therefore be permitted to be dissipative. These equalizers are of the constant resistance type, however, and require approximately twice the number of elements appearing in non-dissipative filters.

pedances of the various elements. On a non-dissipative basis, the reactance elements produce impedances typified by Lp and $\frac{1}{Cp}$. Upon the addition of the uniform dissipation these impedances become $L(p + d)$ and $\frac{1}{C(p + d)}$, where d is a constant representing the value assumed by the equal resistance-reactance ratios at unit ω. It follows that the modification of the complex function corresponding to the addition of the dissipation amounts to the transformation represented by the substitution of $(p + d)$ for p.[94]

It follows from the above principle that the *removal* of uniform dissipation from all the reactance elements of a network will change each corresponding complex function by the transformation represented by the substitution of $(p - d)$ for p. Thus if a network in which the elements are equally dissipative is to be designed to produce a prescribed complex function it is only necessary to design any network producing the "predistorted" function obtained by replacing p by $(p - d)$ in the prescribed function. The prescribed function will then be obtained upon the addition of the required dissipation.

When the insertion loss function of a dissipative reactance network is prescribed, the first step is to find the corresponding insertion voltage ratio exactly as in the design of non-dissipative networks. The predistortion of this voltage ratio and the computation of the corresponding predistorted insertion power ratio is then followed by the design of a pure reactance network, to which the required dissipation is finally added. When the prescribed insertion loss is to be realized exactly, the corresponding voltage ratio of the form $\frac{A + pB}{P}$ yielded by the method of Part I must normally be modified by the reversal of the signs of the poles with negative real parts before predistortion, in order to obtain a predistorted voltage ratio realizable with a pure reactance network.[95] In most actual design problems, however, a more efficient use of elements is obtained by retaining $\frac{A + pB}{P}$ and only approximating the rigorous procedure in a manner to be described later, which

[94] This transformation, which was originally discovered by Bode, is described in Guillemin's "Communication Networks" (18)—Chap. VI, footnote on pp. 245, 246.

[95] It can be shown that this change in the original voltage ratio does not change the corresponding insertion loss but only the insertion phase.

usually leads to a very good approximation to the prescribed insertion loss.

Arrangement of Design Procedure to Meet Physical Conditions

In accordance with Part I, the realization of a predistorted voltage ratio with a pure reactance network calls for the following conditions. First, the predistorted voltage ratio must be expressible in the form $\frac{A + pB}{P}$, in which A, B, and P are even polynomials with real coefficients such that $(A + pB)$ has no roots with positive real parts. Second, the corresponding predistorted power ratio must be positive and no less than $\frac{4R_1R_2}{(R_1 + R_2)^2}$ at all real frequencies. Third, in order to avoid inefficiency in the use of elements it is also best to require that $(A + pB)$ and P shall have no common factors except possibly p.[96]

The usual design procedure is to choose an insertion loss function without regard to dissipation and then to require the dissipation to be small enough to yield a predistorted voltage ratio with roots with negative real parts. Replacing p by $(p - d)$ in the original voltage ratio amounts to adding d to each of the roots and poles. Thus the restriction on the dissipation requires the dissipation constant d to be smaller than the magnitude of the smallest real part of any root of the original voltage ratio, so that the addition of d can not change a negative real part into a positive real part.

In order to meet the limitation on the magnitude of the predistorted power ratio an initially unknown constant factor can be included in the original power ratio. It turns out that this constant also appears in the predistorted power ratio as nothing more than a simple multiplier and can thus easily be chosen in such a way that the predistorted power ratio is no less than $\frac{4R_1R_2}{(R_1 + R_2)^2}$.

If a prescribed insertion loss function is such that the corresponding actual voltage ratio takes the form $\frac{A + pB}{P}$ without use of common

[96] This is necessary for the exclusion of common factors from the expression $\frac{N}{P^2}$ for the corresponding power ratio. Recall that inefficient networks are usually encountered when common factors other than constants or p^2 must be combined with numerator and denominator of the simplest rational fraction expression for $e^{2\sigma}$ in order to obtain the form $\frac{N}{P^2}$.

factors other than p, the predistorted voltage ratio will not take that form.[97] Since this condition requires the finite poles of a voltage ratio to occur in pairs differing only in sign, it cannot be satisfied both before and after the addition of d to each of the poles. Modified insertion loss functions can readily be found such that the predistorted voltage ratios meet this condition rather than the actual voltage ratios. It is normally more satisfactory, however, to start with a prescribed insertion loss function of the previous type and then to realize it only to a good approximation by means of the procedure outlined below.

A power ratio is first chosen which takes the form $\frac{N}{P^2}$ without use of identical roots of N and P^2 except possibly p^2. The theory of Part I is then used to determine the corresponding voltage ratio of the form $\frac{A + pB}{P}$. The predistorted voltage ratio is next obtained by replacing p by $(p - d)$ in the numerator of this function only, the denominator P being left unchanged. A corresponding pure reactance network is then designed, after which the required dissipation is added as before. The voltage ratio actually obtained is a third function derived from $\frac{A + pB}{P}$ by replacing p by $(p + d)$ in the denominator P.

It follows from the fact that P is real at real frequencies that the actual power ratio obtained by the above procedure will approximate the original power ratio $\frac{N}{P^2}$ very closely except near the poles. This is because the first order effect of the substitution of $(p + d)$ for p upon the magnitude of a function of ω is proportional to the derivative of the phase, while the phase of P is zero at real frequencies.[98]

Special Case of Filters

The effectiveness of the design procedure described above is illustrated by the special case in which the original loss function is formed by adding a constant loss to one of the functions described in Part III as appropriate for filter purposes. In this special case, the maximum

[97] Except in the very special case in which P is a constant, i.e., in which all frequencies of infinite loss occur at infinity.

[98] The relation between the derivative of the phase and the effect of the substitution of $(p + d)$ for p upon the magnitude of a function was discovered by Bode, who used it in the derivation of his formulae for the estimation of the effect of adding dissipation to a prescribed network. It is used in exactly the same way by Guillemin (18) in deriving Mayer's solution (10) of the same Problem.

value of the dissipation constant d for which predistortion is physically possible unusually corresponds to an amount of dissipation which would produce a pass band distortion of about 6 decibels if the desired voltage ratio were not predistorted. To avoid practical difficulties it is usually necessary to keep d well below this physical limit, a margin of perhaps 30 to 50 per cent being usually needed.[99]

When d is small enough to satisfy the physical restriction, the predistortion of only the numerator of the voltage ratio frequently leads to a final loss characteristic which differs appreciably from that originally prescribed only by the substitution of rounded finite peaks for the infinite peaks in the attenuation band. It is possible, however, for the loss to be reduced substantially below the original minima between the peaks over a portion of the attenuation band.

Serious reductions in attenuation band losses can be partly compensated for by starting with complex frequencies of infinite loss so chosen that one of each conjugate pair becomes real in the final characteristic. If Tchebycheff pass band parameters are used, for instance, complex m's can be assigned to the reference image parameter filter. It turns out that these m's must be so chosen that the addition of the required dissipation to the reference filter would produce infinite attenuation peaks. This permits known methods to be used in their determination.[100] It is also possible to obtain better approximations to non-dissipative filter characteristics by including so-called compensating resistances in addition to dissipative reactances in the networks producing prescribed insertion losses, but this requires substantial changes in design procedure.

Unequally Dissipative Inductances and Capacities

Suppose now that the previous requirement of uniform dissipation is relaxed to the extent of permitting the capacities to include a different amount of dissipation than the inductances. Except in the design of such circuits as narrow-band filters, the situation immediately becomes much more complicated.[101] The predistorted voltage ratio correspond-

[99] As d approaches the physical limit the constant loss which must be added to the loss function chosen on a non-dissipative basis approaches infinity.

[100] The possibility of using complex m's to obtain infinite attenuation peaks in dissipative image parameter filters was introduced by Bode (4) and is described by Guillemin (18).

[101] A frequency transformation can be used to transform a low-pass filter of equally dissipative inductances and capacities into a band-pass filter of series and parallel resonant circuits resonating at mid-band and with associated series

ing to the removal of all dissipation from the final network can no longer be found directly from the actual voltage ratio. In addition, it now depends upon the assumption of a 4-pole including no resistances other than those produced by the equally dissipative inductances and equally dissipative capacities. It turns out, however, that a design procedure can be obtained by combining the predistortion principle appropriate for equally dissipative inductances and capacities with an additional modification of the theory of Part I.

The simplest procedure is to replace the dissipation constants of the inductances and capacities by constants d_0 and δ representing their average and one half their difference. The typical inductive and capacitive impedances then become $L(p + d_0 + \delta)$ and $\frac{1}{C(p + d_0 - \delta)}$. The predistortion method described above can be used to determine the voltage ratio that would be obtained upon the removal of the dissipation represented by d_0. The design problem is therefore solved if a method can be found for designing a network of impedances of the types $L(p + \delta)$ and $\frac{1}{C(p - \delta)}$ producing a prescribed voltage ratio. Such a design method has been found and turns out to be very similar to that described in Part I for the non-dissipative case, in which δ is zero.

Design of Networks of Oppositely Dissipative Inductances and Capacities

The fact that the above design problem can be solved depends upon the following properties of networks of impedances of the type $L(p + \delta)$ and $\frac{1}{C(p - \delta)}$. In the first place, the elements of the determinant of any network of this type can be expressed in the form.

$$Z_{jk} = \frac{L_{jk}(p^2 - \delta^2) + \frac{1}{C_{jk}}}{p - \delta} \tag{77}$$

It follows that the determinant must be equivalent to the product of $(p - \delta)^{-n}$ and an even polynomial in p, where n is the degree of the determinant. The same statement is also true of any cofactor of the

---

and shunt constant resistances. If the band is narrow, the resistances approximate the total effect of dissipation no matter how it is divided between the inductances and capacities.

determinant provided n is modified to take account of the change in degree. Thus each open- and short-circuit impedance of a network of this particular type is equivalent to the product of $(p - \delta)^{-1}$ and an even rational function of p.[102] In other words, the impedances differ in form from those of non-dissipative networks only by the substitution of the factor $(p - \delta)$ for p.

The important design formulae are those for determining the open- and short-circuit impedances as functions of frequency. After these have been determined the impedances obtained upon the removal of all dissipation can easily be computed. In accordance with (77), all that is necessary is to replace the factor $(p - \delta)^{-1}$ by p^{-1} in each impedance and to replace p^2 by $(p^2 + \delta^2)$ in the even rational function multiplying this factor. Formulae for computing the impedances of the dissipative network are listed below. The derivation will be omitted as in the case of the non-dissipative networks of Part I since the formulae can be checked by simple mesh computations.

Particularly simple formulae are obtained by expressing the insertion voltage ratio in the form

$$\frac{V_{20}}{V_2} = \frac{p}{p - \delta} \frac{A + pB}{P} \tag{78}$$

in which A, B and P are even polynomials as before. The additional polynomials A' and B' are then found exactly as in the determination of non-dissipative networks producing the voltage ratio $\frac{A + pB}{P}$. In other words, A' and B' are determined from the following pair of equations formed by combining (12) and (13) of Part I:

$$\begin{aligned} A^2 - p^2B^2 - \frac{4R_1R_2}{(R_1 + R_2)^2} P^2 &= J'^2(p_1'^2 - p^2) \cdots (p_n'^2 - p^2) \\ A' + pB' &= J'(p_1' - p) \cdots (p_n' - p) \end{aligned} \tag{79}$$

[102] This theorem can also be derived from the following theorem included in Guillemin's "Communication Networks" (18)—Vol. II, Chap. X, p. 445. Suppose each element in a network produces an impedance proportional to $(p + 2\alpha)$ or $\frac{1}{(p + 2\beta)}$. Then any impedance of the network takes the form of the product of $\sqrt{\frac{p + 2\alpha}{p + 2\beta}}$ and an odd rational function of $\sqrt{(p + 2\alpha)(p + 2\beta)}$. This theorem was preceded by a still more general theorem due to Bode which states that if each element of a network produces an individual impedance proportional to Z_1 or Z_2 any impedance of the complete network will be the product of Z_2 and a rational function of $\frac{Z_1}{Z_2}$.

In terms of these polynomials, the impedances of the dissipative network are given by

$$
\begin{aligned}
Z_{O1} &= R_1 \frac{(A - A') + \delta(B + B')}{(p - \delta)(B + B')} \\
Z_{O2} &= R_2 \frac{(A + A') + \delta(B + B')}{(p - \delta)(B + B')} \\
Z_{O12} &= \frac{-2R_1R_2}{(R_1 + R_2)} \frac{P}{(p - \delta)(B + B')} \\
Z_{S1} &= R_1 \frac{p^2(B - B') + \delta^2(B + B') + 2\delta A}{(p - \delta)[(A + A') + \delta(B + B')]} \\
Z_{S2} &= R_2 \frac{p^2(B - B') + \delta^2(B + B') + 2\delta A}{(p - \delta)[(A - A') + \delta(B + B')]} \\
Z_{S12} &= \frac{R_1 + R_2}{2} \left(\frac{p^2(B - B') + \delta^2(B + B') + 2\delta A}{(p - \delta)P} \right)
\end{aligned}
\tag{80}
$$

A very similar alternative set of formulae expressing the short-circuit impedances more simply rather than the open-circuit impedances can be obtained by starting with a voltage ratio derived from (78) by reversing the sign used with δ.

The specific type of voltage ratio function required depends more upon the particular configuration than in the case of non-dissipative networks. A mid-series high pass ladder, for instance, will produce an infinite loss point at $p = +\delta$ while the corresponding mid-shunt ladder will produce one at $p = -\delta$. The exact type of voltage ratio function required can ordinarily be determined by means of conventional circuit analysis. Since the dissipation associated with one type of element is negative, it may be possible for the corresponding power ratio to be less than $\frac{4R_1R_2}{(R_1 + R_2)^2}$ at certain real frequencies. In ordinary design problems, however, it is to be expected that there will be a minimum permissible value somewhere in the near neighborhood of $\frac{4R_1R_2}{(R_1 + R_2)^2}$.

The determination of $(A' + pB')$ by (79) is also more restricted than in the case of non-dissipative networks. This is because it is the multiplicity of solutions for $(A' + pB')$ which leads to the various non-equivalent configurations producing a given voltage ratio, while different configurations which can produce the same voltage ratio on a non-

dissipative basis may no longer have this property after the addition of the dissipation.

Useful information bearing on the choice of B' is obtained by examining the driving-point impedances of a network determined by (80). Simple circuit analysis shows the corresponding reflection coefficients to be

$$\begin{aligned} \frac{R_1 - Z_1}{R_1 + Z_1} &= \frac{p - \delta}{p}\left(\frac{A' + pB'}{A + pB}\right) - \frac{\delta}{p} \\ \frac{R_2 - Z_2}{R_2 + Z_2} &= \frac{p - \delta}{p}\left(\frac{-A' + pB'}{A + pB}\right) - \frac{\delta}{p} \end{aligned} \tag{81}$$

These formulae indicate that at least one of the driving-point impedances will be the negative of the corresponding termination at zero frequency unless $\frac{A'}{pB'}$ and $\frac{A}{pB}$ approach zero at that point and unless $\frac{B'}{B}$ approaches -1. It can be shown that $\frac{A'}{pB'}$ and $\frac{A}{pB}$ will approach zero and $\frac{B'}{B}$ will approach ± 1 provided zero frequency represents an even order pole of the voltage ratio (78), or else a point at which it is finite. The above conditions will then be satisfied by the proper choice of the sign of B', which (79) leaves arbitrary.

Other Types of Uniformly Dissipative Networks

Similar methods can be used in the design of dissipative reactance networks in which each element is equivalent to an inductance or capacity combined with both series and shunt constant resistances, all elements of each kind being equally dissipative at all frequencies. This permits the simulation of the variation of dissipation with frequency encountered in actual elements.

The effect of removing both series and shunt types of dissipation from both kinds of elements in like amounts can be computed by means of a bilinear predistortion transformation. The removal of the proper amount of dissipation of each kind leaves a network in which the capacitive dissipation is the negative of the inductive dissipation. The impedances are then represented by products of $(p - \delta - \gamma p^2)^{-1}$ and even rational functions of p, where δ and γ depend only upon the dissipation constants. Methods similar to those described previously show that the corresponding impedances can be determined by replacing δ by $(\delta + \gamma p^2)$ in (78) and (80).

Theoretically, networks of still more complicated elements can be designed in exactly the same way. The only requirement is the existence of a predistortion transformation producing a predistorted voltage ratio which can be realized with a network of two kinds of elements such that the ratio of their impedances is proportional to an even rational

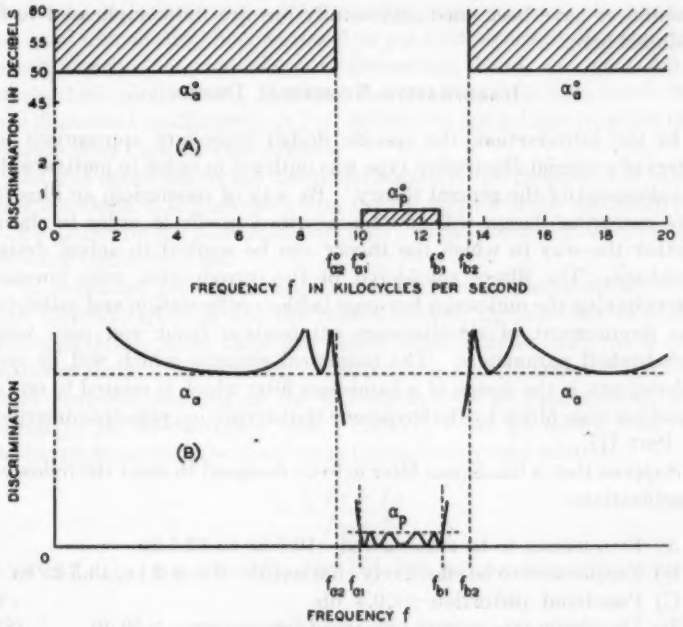


Fig. 24. A) Design specifications—When $0 < f < f_{a2}^\circ$, $f_{a1}^\circ < f < f_{b1}^\circ$, or $f_{b2}^\circ < f < \infty$, the discrimination characteristic must lie in a shaded area. (Discrimination = insertion loss minus an arbitrary constant loss.)

B) Form of discrimination characteristic meeting the design specifications.

function of frequency and such that the inverse of the predistortion transformation leaves them physical impedances. The open- and short-circuit impedances are computed by replacing δ by the proper even rational function. The principal difficulty is the determination of the required form of voltage ratio and the advantageous choice of a specific function of this form.

Concentrated Dissipation

The only non-uniformly dissipative networks which are commonly encountered approximate non-dissiptative networks to which constant resistances have been added at only a few points. Methods similar to those described above do not apply to these networks. Successive approximation methods for correcting for the effect of adding the resistances have been used successfully but are too complicated to be included here.

Illustrative Numerical Design

In the introduction, the specific design procedure appropriate for filters of a special illustrative type was outlined in order to motivate the development of the general theory. By way of conclusion, an illustrative numerical design will now be described briefly in order to clarify further the way in which the theory can be applied to actual design problems. The filters considered in the introduction were low-pass filters having the mid-series low-pass ladder configuration and satisfying the requirement of simultaneous attenuation band and pass band Tchebycheff parameters. The numerical example which will be considered now is the design of a band-pass filter which is related to one of these low-pass filters by the frequency transformation principle described in Part III.

Suppose that a band-pass filter is to be designed to meet the following specifications:

A) Frequencies to be transmitted—10.0 kc to 12.5 kc
B) Frequencies to be effectively eliminated—0 to 9.2 kc, 13.5 kc to ∞
C) Pass-band distortion—<0.4 db
D) Discrimination against unwanted frequencies—>50 db (82)
E) Dissipation compensation to be sufficient to lead to readily realizable elements
F) Final network to be a ladder involving no transformers or coupled coils

Requirements *A* through *D* are indicated graphically in Fig. 24A.

Using the notation indicated in Fig. 24A the critical frequencies f^0_{a1} ; f^0_{a2} , f^0_{b1} , f^0_{b2} satisfy approximately the equation

$$f^0_{a1} f^0_{b1} = f^0_{a2} f^0_{b2} \tag{83}$$

while the discrimination requirement in (82) is the same for both attenuation bands. This amounts to saying that the above loss requirements are symmetrical on a logarithmic frequency scale about the mid-band frequency $\sqrt{f_{a1}^0 f_{b1}^0}$. This symmetry makes it possible to simplify the design procedure by first designing a low-pass filter and then converting it to a band-pass filter by means of a frequency transformation. Due to the uniformity of the discrimination requirement at all attenuation band frequencies, the low-pass filter can be designed by the straightforward method of simultaneous attenuation band and pass band Tchebycheff parameters. The final loss characteristic then takes the form illustrated qualitatively in Fig. 24B for the special case in which the low-pass filter includes two sections.

The first operation in the design procedure is the choice of the arbitrary constants of the low-pass filter theory subject to the specificatons to be satisfied by the final band-pass filter. This choice is guided by the following considerations:

The maximum pass band distortion α_p and the minimum attenuation band discrimination α_a are the same for both the low-pass and the band-pass filters since their loss characteristics differ only as to the frequency scales. The number of sections η is also the same for the two filters provided sections of the 6-element type of configuration are considered in the band-pass case. Finally, it can be shown that the constant k measuring the sharpness of cut-off obeys the relation

$$k = \frac{f_{b1} - f_{a1}}{f_{b2} - f_{a2}} \tag{84}$$

in which f_{a1}, f_{b1}, etc., are defined by Fig. 24B and are themselves related by

$$f_{b1} f_{a1} = f_{b2} f_{a2}. \tag{85}$$

In accordance with Part III, the constants α_p, α_a, k, η of the low-pass theory are not all arbitrary but must be related by (71)—i.e., by

$$\alpha_a = [10 \log_{10}(e^{2\alpha_p} - 1) - 10(2\eta + 1) \log_{10}(q) - 12.04] \text{ db} \tag{86}$$

in which $\log_{10}(q)$ is a quantity which is tabulated against $\sin^{-1}k$ in most elliptic function tables.

A choice of constants which is consistent with (85) and (86) and with the original specifications in (82) is as follows:

$$\begin{aligned} k &= 0.62 \\ \eta &= 2 \\ \alpha_p &= 0.30 \text{ db} \\ \alpha_s &= 52.4 \text{ db} \\ f_{a1} &= 9.96 \text{ kc} \\ f_{b1} &= 12.54 \text{ kc} \\ f_{a2} &= 9.2872 \text{ kc} \\ f_{b2} &= 13.4484 \text{ kc} \end{aligned} \tag{87}$$

These constants fix the final loss characteristic except for changes due to the requirement of dissipative elements.

After the constants of the Tchebycheff theory have been chosen, the corresponding voltage ratio $\frac{A + pB}{P}$ is evaluated. This is easily accomplished by means of the special formulae for the roots and poles which are included in the Tchebycheff theory. If p is used to represent $i\Omega$ rather than $i\omega$, the voltage ratio corresponding to the choice of constants indicated in (87) turns out to be

$$\frac{V_{20}}{V_2} = \frac{A + pB}{P} = \frac{(a_0 + p)(\rho_1^2 + 2a_1 p + p^2)(\rho_2^2 + 2a_2 p + p^2)}{a_0 \rho_1^2 \rho_2^2 (1 + \Omega_1^2 p^2)(1 + \Omega_2^2 p^2)} \tag{88}$$

in which the constants have the following values:

$$\begin{aligned} a_0 &= 0.37766 \\ a_1 &= 0.25943 \\ a_2 &= 0.077333 \\ \Omega_1^2 &= 0.24902 \\ \Omega_2^2 &= 0.57282 \\ \rho_1^2 &= 0.37822 \\ \rho_2^2 &= 0.66141 \end{aligned} \tag{89}$$

The low-pass type of voltage ratio is obtained by defining Ω as $\frac{\omega}{\sqrt{\omega_1 \omega_2}}$

as in equation (60) of Part III. The corresponding band-pass type of voltage ratio is obtained by redefining Ω in accordance with the relation

$$\Omega = \frac{\omega^2 - \omega_{b1}\omega_{a1}}{\omega\sqrt{(\omega_{b1} - \omega_{a1})(\omega_{b2} - \omega_{a2})}} \tag{90}$$

in which ω_{a1}, ω_{b1}, etc. are the values assumed by ω at the frequencies f_{a1}, f_{b1}, etc., defined by Fig. (24B).

The next operation is the predistortion of the voltage ratio (88) in order to compensate for the dissipation required in the elements of the final network. The predistortion is accomplished by replacing p by $(p - d)$ in the numerator of (88). The constant d appearing in this transformation is related to the dissipativeness of the final band-pass filter by

$$d = \frac{\sqrt{f_{a1}f_{b1}}}{\sqrt{(f_{b1} - f_{a1})(f_{b2} - f_{a2})}}(d_L + d_C) \tag{91}$$

in which d_L represents the mean of the resistance-reactance ratios of the coils, evaluated at mid-band frequency, and d_C represents the corresponding mean of the conductance-susceptance ratios of the condensers. In order to obtain a physically realizable design, d_L and d_C must be sufficiently small to render d substantially smaller than a_2 of (89). A suitable pair of values is

$$\begin{aligned} d_L &= 0.0100 \\ d_C &= 0.0025 \end{aligned} \tag{92}$$

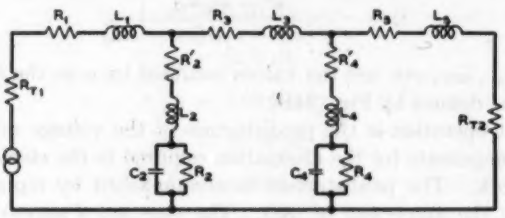
which yields

$$d = 0.04263 \tag{93}$$

In addition to the modification of the voltage ratio (88) by the predistortion transformation described above, a temporarily unknown constant factor is added to represent an added constant loss. The general theories of Parts I and II are then applied to the design of a corresponding pure reactance network, using the frequency variable Ω in place of ω.[103] By choosing the added constant loss and the ratio of terminations in the proper way, a ladder type of network is obtained

[103] If Ω is defined as $\frac{\omega}{\sqrt{\omega_1\omega_2}}$, as for a low-pass filter, it is proportional to ω and can obviously be used in place of it in the formulae of Parts I and II. If it is defined by (90), as for a band-pass filter, the resulting network is related to the low-pass network by the frequency transformation principle.

which includes no transformers or negative impedance branches and which produces a minimum transducer loss.[104] The addition of the



$$R_{T2} = 0.084427\ R_{T1}$$

$$L_1 = 1.1834\ \frac{R_{T1}}{\sqrt{\omega_1\omega_2}} \qquad R_1 = 0.050448\ R_{T1}$$

$$C_2 = \frac{1.8849}{R_{T1}\sqrt{\omega_1\omega_2}} \qquad R_2 = 12.445\ R_{T1}$$

$$L_3 = 0.13211\ \frac{R_{T1}}{\sqrt{\omega_1\omega_2}} \qquad R_3' = 0.0056318\ R_{T1}$$

$$L_3 = 2.3227\ \frac{R_{T1}}{\sqrt{\omega_1\omega_2}} \qquad R_3 = 0.099017\ R_{T1}$$

$$C_4 = \frac{1.7650}{R_{T1}\sqrt{\omega_1\omega_2}} \qquad R_4 = 13.290\ R_{T1}$$

$$L_4 = 0.32454\ \frac{R_{T1}}{\sqrt{\omega_1\omega_2}} \qquad R_4' = 0.013835\ R_{T1}$$

$$L_5 = 0.85255\ \frac{R_{T1}}{\sqrt{\omega_1\omega_2}} \qquad R_5 = 0.036344\ R_{T1}$$

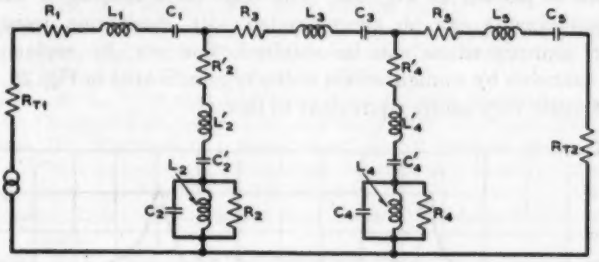
Provided R_{T1} is in ohms and ω_1, ω_2 are in radians per second:
Inductances are in henries
Capacities are in farads
Resistances are in ohms

Fig. 25. Preliminary low-pass filter

proper resistances to this network, corresponding to the substitution of $(p + d)$ for p in the impedance functions representing the different

[104] Recall that transducer loss was defined in the introduction as the difference in level between the received power and the maximum power obtainable from the generator with any passive network.

branches, leads to a good approximation to the loss determined by (88) plus an added constant loss.



$$R_{T2} = 0.084427\ R_{T1}$$

$L_1 = 0.057481\ R_{T1}$	$C_1 = \dfrac{3.5282}{R_{T1}}$	$R_1 = 0.050448\ R_{T1}$
$L_2 = 0.0022151\ R_{T1}$	$C_2 = \dfrac{91.556}{R_{T1}}$	$R_2 = 12.445\ R_{T1}$
$L_2' = 0.0064170\ R_{T1}$	$C_2' = \dfrac{31.605}{R_{T1}}$	$R_2' = 0.0056318\ R_{T1}$
$L_3 = 0.11282\ R_{T1}$	$C_3 = \dfrac{1.7976}{R_{T1}}$	$R_3 = 0.099017\ R_{T1}$
$L_4 = 0.0023656\ R_{T1}$	$C_4 = \dfrac{85.732}{R_{T1}}$	$R_4 = 13.290\ R_{T1}$
$L_4' = 0.015764\ R_{T1}$	$C_4' = \dfrac{12.865}{R_{T1}}$	$R_4' = 0.013835\ R_{T1}$
$L_5 = 0.041411\ R_{T1}$	$C_5 = \dfrac{4.8074}{R_{T1}}$	$R_5 = 0.036344\ R_{T1}$

Provided R_{T1} is in ohms:
Inductances are in millihenries
Capacities are in microfarads
Resistances are in ohms

Fig. 26. Band-pass filter derived from the preliminary low-pass filter by a frequency transformation

If a mid-series type of ladder configuration is assumed and if Ω is defined as $\dfrac{\omega}{\sqrt{\omega_1 \omega_2}}$, the above procedure leads to the low-pass filter indicated in Fig. 25. Transforming this network by redefining Ω in

terms of (90), with f_{a1}, f_{a2}, etc., fixed by (87), yields the band-pass design indicated in Fig. 26. The computed loss characteristic of this network is plotted in Fig. 27. The resistances associated with the resonant circuits can be approximated with dissipation resistances. Better approximations can be obtained, however, by replacing the shunt branches by configurations of the type indicated in Fig. 28, which can be made very nearly equivalent to them.

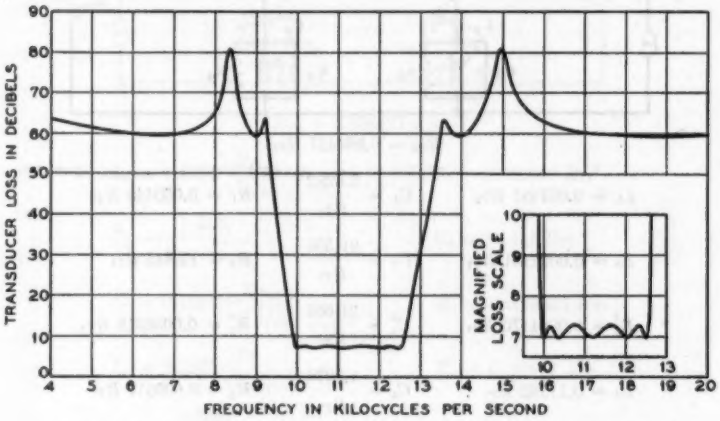


Fig. 27. Transducer loss of the network of Fig. 26 (i.e., insertion loss plus reflection loss corresponding to inequality of terminations)

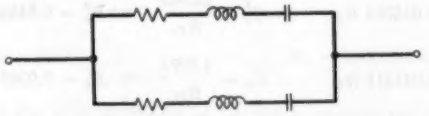


Fig. 28. Alternative shunt branch configuration

BIBLIOGRAPHY

1. Pupin, M. I.: "Wave Propagation over Non-Uniform Cables and Long Distance Air-Lines," *Trans. A. I. E. E.*, *17*, 445–507 (1900).
2. Campbell, G. A.: "On Loaded Lines in Telephonic Transmission," *Phil. Mag.*, *5*, 313–330 (1903).
3. Zobel, O. J.: "Theory and Design of Uniform and Composite Electric Wave Filters," *Bell System Tech. Journal*, *2*, 1–46 (1923), and "Transmission Characteristics of Electric Wave Filters," *Bell System Tech. Journal*, *3*, 567–620 (1924).

4. Bode, H. W.: "A General Theory of Electric Wave Filters," *Journal of Math. and Physics, 13,* 275–362 (1934).
5. Cauer, W.: "Ein Reaktanztheorem," *Preuss. Akad. d. Wissenschaften, Phys.-Math. Kl., Sitzber., Nos. 30–32,* 673–681 (1931).
6. Gewertz, C. M.: "Synthesis of a Finite Four-Terminal Network from its Prescribed Driving-Point Functions and Transfer Function," *Journal of Math. and Physics, 12,* 1–257 (1933).
7. Norton, E. L.: "Constant Resistance Networks with Applications to Filter Groups," *Bell System Tech. Journal, 16,* 178–193 (April, 1937).
8. Brune, O.: "Synthesis of a Finite Two-Terminal Network Whose Driving-Point Impedance is a Prescribed Function of Frequency," *Journal of Math. and Physics., 10,* 191–235 (1931).
9. Dietzold, R. L.: "A Mechanical Root Finder," *Bell Labs. Record, 16,* Dec., 1937.
10. Mayer, H. F.: "Ueber die Daempfung von Siebketten in Durchlaissigkeitsbereich," *E. N. T., 2,* 335–338 (1925).
11. Bode, H. W. and Dietzold, R. L.: "Ideal Wave Filters," *Bell System Tech. Journal, 14,* 215–252 (1935).
12. Fry, T. C.: "The Use of Continued Fractions in the Design of Electrical Networks," *Am. Math. Soc., Bull., 35,* 463–498 (1929).
13. Scott, R. F. and Mathews, G. B.: "The Theory of Determinants," Second Edition, 1904.
14. Norton, E. L.: U. S. Patent No. 1681554.
15. Campbell, G. A.: "Physical Theory of the Electric Wave-Filter," *Bell System Tech. Journal, 1,* 1–32 (1922).
16. Cauer, W.: "Ein Interpolationsproblem mit Funktionen mit Positiven Realteil," *Mathematische Zeitschrift, 38,* 1–44 (1933).
17. Shea, T. E.: "Transmission Networks and Wave Filters," D. Van Nostrand Co., New York, 1929—p. 253, Fig. 138.
18. Guillemin, E. A.: "Communication Networks," John Wiley and Sons, New York, N. Y., 1935—Vol. II.
19. Cayley, A.: "An Elementary Treatise on Elliptic Functions," second edition, G. Bell and Sons, London, 1895.
20. Jacobi, C. G. J.: "Fundamenta Nova Theoriae Functionum Ellipticarum," Königsberg, 1829.
21. Silberstein, L.: "Synopsis of Applicable Mathematics," G. Bell and Sons, London, 1923—Section on Elliptic Functions, pp. 152–170.
22. Bode, H. W.: U. S. Patents Nos. 1955788, 2002216, 2029014.